



**Facultad de Ingeniería y Computación
Escuela Profesional de Ingeniería Electrónica y de
Telecomunicaciones**

**“Estudio Comparativo para la mejora de
Imágenes de Documentos con Diferentes
Familias Tipográficas Usando la Transformada
Wavelet, Curvelet y NMF”**

Presentado por:

Antonio Emmanuel Lloque Juárez

Para Optar por el Título Profesional de:

Ingeniero de Telecomunicaciones

Orientador: Dr. Jimmy Ludeña Choez

Arequipa, Enero de 2018

**PROGRAMA PROFESIONAL DE INGENIERÍA DE
TELECOMUNICACIONES**

**Estudio Comparativo para la mejora de Imágenes de
Documentos con Diferentes Familias Tipográficas Usando la
Transformada Wavelet, Curvelet y NMF**

Autor: Antonio Emmanuel Lloque Juárez

Enero, 2018

Asesor: Dr. Jimmy Ludeña Choez

“Nuestra recompensa se encuentra en el esfuerzo y no en el resultado. Un
esfuerzo total es una victoria completa.”

Mahatma Gandhi, 1869-1948

Índice general

Abstract	5
Resumen	7
1. Introducción	9
1.1. Motivación y Contexto	10
1.2. Planteamiento del problema	10
1.2.1. Objetivo general	11
1.2.2. Objetivos específicos	11
1.3. Metodología	11
1.4. Consideraciones Complementarias	12
1.4.1. Recursos y Materiales	12
2. Marco Teórico	13
2.1. Representación de imágenes digitales	13
2.2. Ruido en imágenes digitales	14
2.2.1. Ruido gaussiano	14
2.2.2. Ruido impulsivo	15
2.2.3. Ruido multiplicativo	16
2.3. Medidas de Evaluación de calidad de imagen	17
2.3.1. PSNR	18
2.3.2. MSSIM	19
2.3.3. MOS	21
2.4. Tipografía	22
2.4.1. Tipografía con serifa	22
2.4.2. Tipografía sin serifa	23
2.4.3. Tipografía máquina de escribir	23
2.5. Métodos de eliminación de ruido	24
2.5.1. Transformada Wavelet	25
2.5.2. Transformada Curvelet	29
2.6. Factorización de Matrices no-Negativa	35
2.6.1. Definición	35
2.6.2. Modelo Básico	35
2.6.3. Funciones de coste	37
2.6.4. Reglas de aprendizaje multiplicativo	37
2.6.5. Restricciones adicionales	38

2.6.6. Aplicaciones	39
3. Estado del Arte	43
4. Metodología de Solución y Diseño	49
4.1. Método basado en wavelet	49
4.2. Método basado en curvelet	50
4.3. Método basado en NMF	52
4.3.1. Entrenamiento	54
4.3.2. Eliminación de Ruido	55
5. Experimentación y Análisis	59
5.1. Mediciones basadas en PSNR y MSSIM	59
5.2. Mediciones basadas en MOS	61
5.2.1. Diseño experimental para la medida MOS	63
5.2.2. Protocolo experimental para la obtención de la medida MOS	64
5.2.3. Cálculo MOS	65
5.3. Resultados experimentales	65
5.3.1. Tipografía con serifa	66
5.3.2. Tipografía sin serifa	71
5.3.3. Tipografía máquina de escribir	76
5.3.4. Análisis comparativo	81
6. Conclusiones y trabajos futuros	87
6.1. Conclusiones	87
6.2. Trabajos futuros	88
Acknowledgements	89
Bibliografía	91
Nomenclatura	99

Índice de figuras

2.1. Proceso de digitalización de una imagen: a) Muestreo y b) Cuantización. [GW02].	13
2.2. a) Imagen original sin ruido y b) Imagen contaminada con ruido gaussiano de media cero y varianza 0.04.	14
2.3. a) Imagen original sin ruido y b) Imagen contaminada con ruido impulsivo de media cero y varianza 0.04.	15
2.4. Imagen contaminada con ruido speckle.	17
2.5. Ejemplo de letra con serifa.	23
2.6. Ejemplo de letra sin serifa.	24
2.7. Ejemplo de letra del tipo maquina de escribir.	24
2.8. Esquema básico de un proceso de eliminación de ruido.	25
2.9. Esquema de la Transformada Wavelet.	27
2.10. Descomposición piramidal de coeficientes wavelet a tres niveles [Bal07].	28
2.11. a) Imagen original de Barbara (izquierda), b) Descomposicion a dos niveles (derecha) [Bal07].	29
2.12. Análisis y síntesis en 2D de un solo nivel [Can10].	29
2.13. Embaldosado del dominio de la frecuencia en cuñas para la construcción de curvelets y el área sombreada representa una cuña (wedge) genérica [CDDY05].	31
2.14. Plano seccionado de la Transformada Curvelet con soporte pseudopolar parabólico en el plano de frecuencia [CDDY05].	32
2.15. Diagrama de proceso de FDCT [SMEG12].	34
2.16. La multiplicación entre W y H nos dan una reconstrucción aproximada de V	36
4.1. Diagrama de bloques para el proceso de eliminación de ruido en imágenes basado en Transformada Wavelet [PVL12].	51
4.2. Diagrama de bloques del proceso de eliminación de ruido en imágenes usando el método basado en Transformada Curvelet.	52
4.3. Representación básica sobre una imagen ruidosa al ser descompuesta por NMF [LG15].	54
4.4. Diagrama de bloques del proceso de eliminación de ruido usando el método basado en NMF con control de restricción de dispersidad. .	57

5.1. Seis diferentes tipografías de la base de datos: a) Letra con serifa itálica, b) Letra con serifa, c) Letra sin serifa itálica, d) Letra sin serifa, e) Letra máquina de escribir itálica y f) Letra máquina de escribir.	60
5.2. Ejemplo de secuencia, conformada por una muestra de referencia y tres de prueba, cada una procesada por diferente método de eliminación de ruido. Todas las opciones bajo las mismas condiciones. . . .	62
5.3. Una sección de las hojas de calificación.	65
5.4. Resultados PSNR para tipografías con serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	68
5.5. Resultados MSSIM para tipografías con serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	69
5.6. Resultados MOS para tipografías con serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	70
5.7. Resultados PSNR para tipografías sin serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	73
5.8. Resultados MSSIM para tipografías sin serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	74
5.9. Resultados MOS para tipografías sin serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	75
5.10. Resultados PSNR para tipografías maquina de escribir contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	78
5.11. Resultados MSSIM para tipografías maquina de escribir contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	79
5.12. Resultados MOS para tipografías maquina de escribir contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.	80
5.13. Ejemplos de imágenes de documentos tras el proceso de eliminación de ruido basado en NMF con tipografía: a) con serifa y b) sin serifa. Las imágenes fueron contaminadas con ruido gaussiano de media cero y varianza 0.05.	83
5.14. Ejemplos de imágenes de documentos de tipografía con serifa tras pasar por los métodos de eliminación de ruido basados en : a) wavelet, b) curvelet, c) NMF. Las imágenes fueron contaminadas con ruido impulsivo de media cero y varianza 0.06.	84
5.15. Ejemplos de imágenes de documentos tras pasar por el método de eliminación de ruido basado en wavelet con tipografía: a) con serifa, b) sin serifa, c) maquina de escribir y d) imagen ruidosa . Las imágenes fueron contaminadas con ruido gaussiano de media cero y varianza 0.09.	85

Índice de cuadros

2.1. La escala más popular de Calificación de Categoría Absoluta.	21
5.1. Cantidad de imágenes a prueba por tipografía.	61
5.2. Escala de medición usada para la medición MOS.	63
5.3. Tabla resumen de los resultados mostrados con la tipografía con serifa.	67
5.4. Tabla resumen de los resultados mostrados con la tipografía sin serifa.	72
5.5. Tabla resumen de los resultados mostrados con la tipografía máquina de escribir.	77
5.6. Tablas de promedio de las medidas de evaluación: a) PSNR, b) MS- SIM y c) MOS, considerando las distintas varianzas de ruido que contamianan las imágenes con tipografía con serifa.	81
5.7. Tablas de promedio de las medidas de evaluación: a) PSNR, b) MS- SIM y c) MOS, considerando las distintas varianzas de ruido que contamianan las imágenes con tipografía sin serifa.	82
5.8. Tablas de promedio de las medidas de evaluación: a) PSNR, b) MS- SIM y c) MOS, considerando las distintas varianzas de ruido que contamianan las imágenes con tipografía maquina de escribir.	82

Abstract

Nowadays, the process of digitizing documents is very important; however, the presence of digital noise generates a big problem that mainly affects the visual quality of the text, preventing the comfort of the people at the time of reading. For this reason, the main objective of this thesis is to carry out a comparative study between two conventional methods of noise elimination based on the Wavelet Transform and the Curvelet Transform with the NMF-based method. NMF allows you to successfully extract features from an image and it presents robustness in the presence of noise due to its dispersive representation of data. However, NMF does not consider explicit control of the sparse representation of the image in its base vectors and activation or gain coefficients. In this work, the Kullback-Leibler cost function was applied with regulation parameters in order to control the dispersion degree of the NMF algorithm. In addition, an Otsu binarization and spectral mask was applied as post-processing.

In order to evaluate the performance of the aforementioned noise elimination methods, objective (PSNR and MSSIM) and subjective (MOS) quality assessment measures will be used using a database of document images with three different typographic families (with serifa, without serif and typewriter).

The experimental results show that depend on typography and type and noise level, some of the disposal methods used present a better performance than the other. For example, in the case of the method based on the Wavelet Transform, it has better results in images of the document contaminated with impulsive noise when measured with subjective measure MOS, in the case of the method based on the Curvelet Transform, it is the one that in general presents better performance in the process of elimination of noise. While, in the case of the NMF-based method, it presents an appropriate performance of PSNR and MSSIM by reducing gaussian, impulsive and multiplicative noise with typographies with and without serifs, surpassing the Wavelet Transform-based method and according to the results based on the MOS measure reduces their differences with the method based on Curvelet Transform.

Resumen

En la actualidad el proceso digitalización de documentos es muy importante; sin embargo la presencia de ruido digital genera un gran problema que afecta principalmente a la calidad visual del texto impidiendo la comodidad de las personas al momento de leer. Por este motivo, el objetivo principal de este trabajo de tesis consiste en realizar un estudio comparativo entre dos métodos convencionales de eliminación de ruido basado en la Transformada Wavelet y la Transformada Curvelet con el método basado en NMF. NMF permite extraer exitosamente características de una imagen y presenta robustez frente a la presencia de ruido debido a su capacidad de representación dispersa de los datos. Sin embargo, NMF no considera un control explícito de la representación dispersa de la imagen en sus vectores base y coeficientes de activación o ganancia. En este trabajo se aplicó la función de coste de Kullback-Leibler con parámetros de regulación para poder controlar el grado de dispersión del algoritmo NMF. Además, se aplicó un post-procesamiento de binarización Otsu y máscara espectral.

Con la finalidad de evaluar el rendimiento de los métodos de eliminación de ruido mencionados anteriormente se usarán medidas de evaluación de calidad objetivas (PSNR y MSSIM) y subjetivas (MOS) usando una base de datos de imágenes de documentos con tres diferentes familias tipográficas (con serifa, sin serifa y máquina de escribir).

Los resultados experimentales demuestran que dependiendo de la tipografía y tipo y nivel de ruido, algunos de los métodos de eliminación usados presentan un mejor rendimiento que el otro. Por ejemplo, para el caso del método basado en la Transformada Wavelet presenta mejores resultados en imágenes de documento contaminadas con ruido impulsivo al ser medidas con la medida subjetiva MOS, en el caso del método basado en la Transformada Curvelet es el que en general presenta mejor rendimiento en el proceso de eliminación de ruido. Mientras que en el caso del método basado en NMF presenta un rendimiento apropiado considerando las medidas objetivas PSNR y MSSIM al reducir ruido gaussiano, impulsivo y multiplicativo con tipográficas con y sin serifa superando al método basado en Transformada Wavelet; y de acuerdo a los resultados basados en la medida MOS, reduce sus diferencias con el método basado en la Transformada Curvelet.

1 Introducción

Vivimos en un mundo donde la tendencia a la digitalización se va incrementando a pasos agigantados. Aun así, gran parte de la documentación que manejamos la conservamos en papel. Desafortunadamente, su frágil y simple estructura lo hacen altamente deteriorable no solo con el paso del tiempo sino también por otros infortunios que la vida misma nos trae día a día. Es esto lo que nos conlleva a la siguiente pregunta ¿qué pasaría si esta documentación desapareciera por no haber pasado por un buen proceso de digitalización? Simplemente, esto significaría una pérdida total en información, tiempo, recursos humanos y dinero. Sería aún más grave para aquellas empresas que no lograron digitalizar sus documentos a tiempo.

Sin embargo de nada sirve que la información digitalizada sea ilegible. Es por ello que durante los últimos años se ha desarrollado múltiples métodos y algoritmos que permiten la eliminación de ruido en imágenes digitales y en especial en aquellos que contienen texto. Para lograr una buena calidad en nuestros documentos digitales no basta con simplemente tomarles una foto o escanearlos, ya que durante el proceso de digitalización se puede introducir ruido que afecta a la visibilidad del texto. Y para colmo de males, puede que el documento venga de antemano ya en mal estado físico, como por ejemplo, manchas, mala impresión, papel arrugado, etc.

Por todo ello es que la eliminación de ruido en imágenes en documentos se está volviendo en una tarea importante en estos días para mejorar el estilo de vida digital que ya tenemos. Conociendo estas circunstancias es que proponemos un estudio comparativo entre dos métodos convencionales de eliminación de ruido en imágenes basado en la Transformada Wavelet y Transformada Curvelet con el método basado en la Factorización de Matrices No-Negativa (NMF, Non-Negative Matrix Factorization) con control de la restricción de dispersión. NMF tiene la propiedad de poder descomponer una matriz en otras dos matrices de componentes no negativos que junto con el control de la restricción de dispersidad pueden extraer las características base de la imagen y de esta manera eliminar el ruido.

El resto del trabajo está dividido de la siguiente manera. En la Sección 2, se muestra el Marco Teórico donde se describe la teoría básica sobre herramientas matemáticas usadas para atacar el problema de la eliminación de ruido en imágenes. En la Sección 3, se muestra los trabajos relacionados que ayudaron con la realización de este proyecto. En la Sección 4, se explica los métodos basados en wavelet, curvelet y el propuesto basado en el algoritmo NMF. En la Sección 5, se exhiben los resultados experimentales, con sus respectivos análisis. Finalmente, en la Sección 6 se dan a conocer las conclusiones finales.

1.1. Motivación y Contexto

El simple hecho de hablar de digitalización de documentos ya trae consigo muchos beneficios, como por ejemplo: en lo ambientalista, se estaría reduciendo el consumo de papel y de por si la tala de árboles; en lo empresarial, traería un mejor manejo, almacenamiento y distribución de su documentación; en lo cultural, guardaríamos para siempre el legado e historia de la humanidad.

Sin embargo, es inevitable que las imágenes de documentos sufran de algún tipo de distorsión o ruido sobre todo en el proceso de la digitalización, por lo que es fundamental aplicar un proceso de eliminación de ruido en imágenes. En este trabajo de tesis se realizará un estudio comparativo entre los métodos convencionales de eliminación de ruido en imágenes digitales basado en la Transformada Wavelet y Curvelet y el método basado en NMF con control de la restricción de dispersión.

Qué mayor motivación que ser parte activa de la solución al evaluar métodos de eliminación de ruido en imágenes de documentos para mejorar la calidad de las imágenes, facilitando la comodidad de lectura en personas e incluso mejorando el reconocimiento de caracteres en sistemas OCR (Reconocimiento Óptico de Caracteres, Optical Character Recognition).

1.2. Planteamiento del problema

La principal fuente de ruido digital en imágenes surge durante los procesos de adquisición (digitalización), transmisión o procesamiento de imágenes. El rendimiento de los sensores de imágenes se ven afectados por distintos motivos durante la adquisición de imágenes. Por ejemplo: las condiciones ambientales, el nivel de luminosidad (ambientes con baja luminosidad requieren de una amplificación de alta ganancia), temperatura de sensores (a una mayor temperatura se amplifica más el ruido). Mientras que en su transmisión se presentan los siguientes inconvenientes: interferencia en el canal de transmisión por destellos de luz, disturbios atmosféricos, etc. con lo que se obtienen bits de información erróneos. Otras causas son el mal funcionamiento de la tarjeta gráfica, el software de edición de imágenes, defectos de algunos mecanismos del equipo electrónico (escáneres y cámaras digitales) con el que se obtiene la imagen.

El problema radica en el ruido que se adhiere a las imágenes de documentos digitales reduciendo la calidad visual del texto dificultando la lectura en las personas y provocando un pobre reconocimiento de caracteres en diversos sistemas operativos. Es por ello la importancia de diseñar sistemas de eliminación de ruido.

1.2.1. Objetivo general

Realizar un estudio comparativo entre dos métodos convencionales de eliminación de ruido en imágenes documentos basado en la Transformada Wavelet y Transformada Curvelet con el método basado en NMF con control de restricciones de dispersividad, a través de medidas de evaluación objetiva como la Relación Señal a Ruido Pico (PSNR, Peak Signal to Noise Ratio) y Similitud Estructural Media (MSSIM, Mean Structural Similarity); y evaluación subjetiva como la Puntuación de Opinión Media (MOS, Mean Opinion Score) usando una base de datos de imágenes de documentos que contiene tres familias tipográficas (con serifa, sin serifa y máquina de escribir).

1.2.2. Objetivos específicos

1. Implementar el sistema de eliminación de ruido en imágenes de documentos basado en la Transformada Wavelet.
2. Implementar el sistema de eliminación de ruido en imágenes de documentos basado en la Transformada Curvelet.
3. Implementar el sistema de eliminación de ruido en imágenes de documentos basado en NMF.
4. Comparar los tres sistemas de eliminación de ruido en imágenes de documentos a través de medidas de evaluación objetivas y subjetiva usando una base de datos de imágenes de documentos con tres familias tipográficas.

1.3. Metodología

La metodología usada en este trabajo es la siguiente: Primero, se realiza una revisión bibliográfica básica para conocer las ventajas, desventajas, características, etc. que nos ayude a plantear un esquema adecuado para la eliminación de ruido en imágenes de documentos basados en NMF. Luego, se realizó una investigación de otros métodos de eliminación de ruido en imágenes como lo son los basados en la Transformada Wavelet y Transformada Curvelet. A partir de ahí se hará un análisis y comparación, a través del uso de medidas de evaluación de calidad objetivas y subjetiva, entre todos los métodos estudiados.

Se implementará, experimentará y obtendrá los resultados concluyentes de los diferentes métodos a través de medidas de evaluación de calidad objetiva (PSNR y MSSIM) usando el software MATLAB. Otra gran herramienta para la simulación es la toolbox de las curvelets, brindado por CurveLab que se encuentra disponible en [CDDY15], ya que contiene secuencias de código de MATLAB de la Transformada Curvelet Rapida Discreta (FDCT, Fast Discrete Curvelet Transform) vía wrapping.

En cuanto a la experimentación subjetiva para evaluar la calidad de las imágenes se uso la Puntuación de Opinión Media (MOS), se ha tomado en cuenta las recomendaciones de la ITU-T P.910 (Sector de Normalización en las Telecomunicaciones). Finalmente, se mide la calidad de las tipografías analizadas tras pasar por un proceso de eliminación de ruido mediante la comparación entre las propuestas existentes basadas en wavelet y curvelet con la propuesta que se presenta en este trabajo basada en NMF.

1.4. Consideraciones Complementarias

1.4.1. Recursos y Materiales

- Hardware: laptop INTEL Core i5-2450M, procesador de 2.5 Ghz y una memoria RAM de 8 GB.
- Software: MATLAB R2013a
- Toolbox CurveLab.
- Base de datos de imágenes de documentos obtenidos en Kaggle [Inc17].

2 Marco Teórico

2.1. Representación de imágenes digitales

Una imagen se representa mediante una función continua bidimensional $f(x, y)$ [GW02, Mar13, YGvV07], donde x e y representan las coordenadas espaciales de la imagen, y la amplitud de la función f es el nivel de luminosidad o nivel de gris en cualquier punto (x, y) . La unidad mínima fundamental de la imagen recibe el nombre de pixel (Figura 2.1).

La adquisición de la imagen es el primer paso para su procesamiento digital y este puede ser realizado comúnmente por escáneres o cámaras digitales. Para obtener la imagen digital se necesita convertir los datos continuos captados por los sensores a su forma digital mediante procesos de muestreo (discretización espacial) y cuantización (discretización de amplitud). En la Figura 2.1 se observa el proceso de digitalización (muestreo y cuantización) donde podemos observar como se afecta la resolución y calidad de la imagen digital.

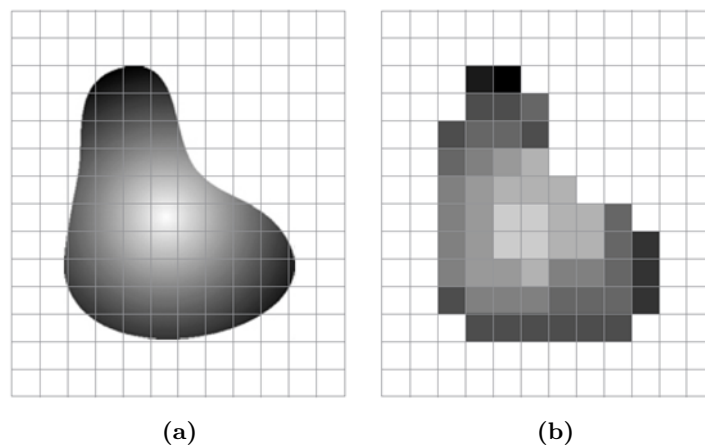


Figura 2.1: Proceso de digitalización de una imagen: a) Muestreo y b) Cuantización. [GW02].

Para el proceso de cuantización se utiliza una cantidad de bits. Para el caso de una imagen binaria toma un solo bit, el pixel será negro cuando su valor es 0 o blanco cuando su valor es 1. En cambio, si la imagen es a escala de grises contiene 8 bits por pixel, lo que significa que tiene 256 (0 a 255) niveles de grises, donde 0 representa al

negro y 255 al blanco. En el caso de las imágenes a color este contiene 3 bandas cada una de diferente color, es decir, cada pixel es expresado en 3 componentes (azul, verde y rojo), mejor conocido como el modelo RGB; siendo cada pixel representado por 24 bits [Mar13, GW02].

2.2. Ruido en imágenes digitales

El ruido según [Pol07] se define como toda perturbación electromagnética que afecta a un circuito digital, toda señal parásita no propia del comportamiento del circuito que puede producir errores. Pero en lo que se refiere específicamente a imágenes, el ruido es una variación aleatoria del color o brillo que afecta a los pixeles produciendo efectos de distorsión que afectan en gran medida la nitidez y calidad de la imagen.

A continuación, se nombraran los tipos de ruidos más comunes y conocidos en el procesamiento digital de imágenes [PVL12, GS13, BJ15, Cat16, GW02].

2.2.1. Ruido gaussiano

Este tipo de ruido afecta a cada uno de los pixeles que componen la imagen cambiando su valor siguiendo una distribución normal o gaussiana. Un ejemplo de cómo afecta este tipo de ruido a una imagen se observa en Figura 2.2

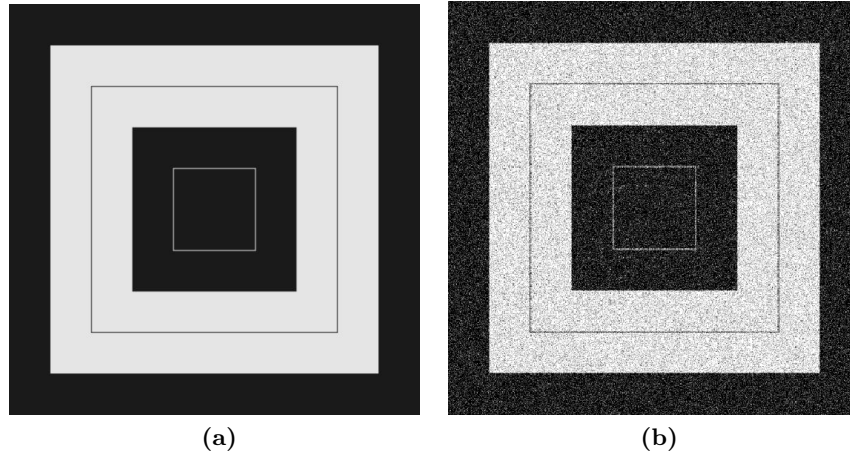


Figura 2.2: a) Imagen original sin ruido y b) Imagen contaminada con ruido gaussiano de media cero y varianza 0.04.

La Función de Densidad de Probabilidad (PDF, Probability Density Function) p_g de una variable gaussiana aleatoria z está dada por

$$p_g(z) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{(z - \mu)^2}{2\sigma^2}\right) \quad (2.1)$$

donde z representa el nivel de gris, μ representa la media de z y σ representa la desviación estándar del ruido. La desviación estándar al cuadrado (σ^2) se llama la varianza de ruido.

El ruido gaussiano aparece en la etapa de adquisición de imágenes digitales causado por la pobre iluminación y/o altas temperaturas; y en su transmisión como por ejemplo ruido en circuitos electrónicos. La solución más común a la contaminación de ruido gaussiano es usar filtros espaciales pero lamentablemente suaviza demasiado las imágenes, lo que provoca que se produzca desenfoque en los bordes y detalles de la imagen a escala fina ya que estos corresponden a frecuencias altas. Incluso una imagen con solo ruido gaussiano al ser binarizada puede presentar también ruido impulsivo. Esta clase de ruido corrompe la calidad de la imagen y causa el pobre reconocimiento de caracteres.

2.2.2. Ruido impulsivo

El ruido impulsivo también conocido como ruido de sal y pimienta (Figura 2.3) se caracteriza porque presenta cambios bruscos en la intensidad de la imagen.

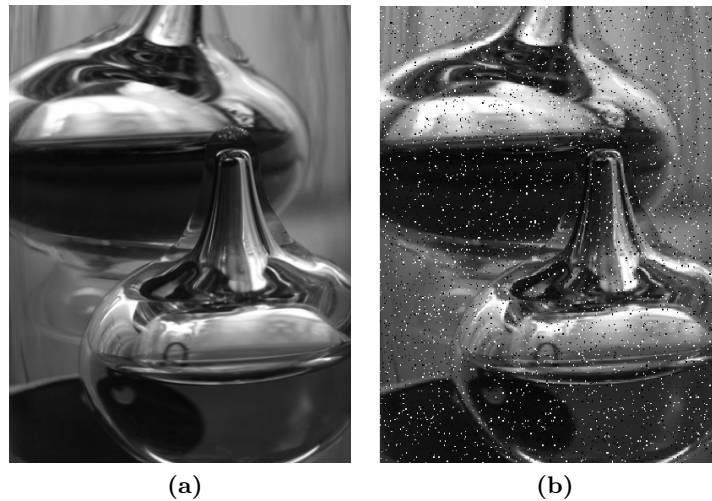


Figura 2.3: a) Imagen original sin ruido y b) Imagen contaminada con ruido impulsivo de media cero y varianza 0.04.

La PDF de un modelo de ruido impulsivo bipolar presenta solo dos valores a y b cuya probabilidad es menor a 0.1 y toman el valor mínimo o máximo, su función de probabilidad está dado por

$$p_i(z) = \begin{cases} P_a & z = a \\ P_b & z = b \\ 0 & \text{otro caso} \end{cases} \quad (2.2)$$

Si $b > a$, el nivel de gris de b aparece como un punto de luz (como la sal) en la imagen, y de modo contrario si $a > b$ entonces el nivel de gris de a será un punto oscuro (como la pimienta). Si alguno de los dos P_a o P_b es igual a cero entonces la PDF es llamado unipolar. Además, la corrupción impulsiva es más grande comparada con la intensidad de la señal.

Las causas más recurrentes que producen este tipo de ruido son las manchas de polvo dentro de las ópticas de las cámaras, sensores defectuosos que realizan una captura de imagen errónea, errores en la conversión analógico-digital, errores de bit en la transmisión a través de un canal, errores de sincronización o ubicación defectuosa de memoria en el proceso de digitalización, etc.

Además, contamina un documento mientras es escaneado o después de la binarización de una imagen de documento escaneada. Un síntoma negativo que produce el ruido impulsivo es que causa el falso reconocimiento de caracteres.

2.2.3. Ruido multiplicativo

En el procesamiento de señales, el término ruido multiplicativo se refiere a una señal aleatoria no deseada que se multiplica con una señal relevante durante la captura, transmisión u otro procesamiento. Según lo explicado en [TM13], dentro de los tipos de ruido multiplicativo se encuentra el ruido speckle que es un tipo de contaminación granular propia de las microondas, siendo producidas por la interferencia de la señal con objetos de dimensiones comparables a la longitud de onda que aparece en imágenes con iluminación coherente y también se origina debido a la interferencia de la onda de retorno en la abertura del transductor. Este tipo de ruido está formado por la suma de ecos de señales cada una prediciendo una cantidad aleatoria de energía.

En [BF10] indican que este ruido se multiplica a la imagen original y su función de densidad de probabilidad sigue una distribución gamma. Donde, g es el nivel de gris y α la varianza.

$$p_m(z) = \frac{g^{\alpha-1} e^{-g/a}}{(\alpha-1)! a^{-\alpha}} \quad (2.3)$$

En lo que respecta a fotografía digital, el ruido multiplicativo afecta a las imágenes que presentan: sombras debido a ondulaciones en la superficie de los objetos de la imagen, sombras proyectadas por objetos complejos, manchas oscuras causadas por

el polvo en el lente o en el sensor de imagen y variaciones en la ganancia de elementos individuales de la matriz de sensores de imagen [PP10].

Se puede agregar que es un tipo de ruido que degrada los detalles finos y la definición de los bordes, y limita la resolución de contraste.



Figura 2.4: Imagen contaminada con ruido speckle.

2.3. Medidas de Evaluación de calidad de imagen

El objetivo de toda evaluación de calidad de imágenes es obtener un resultado concreto que refleje el proceso del experimento. Para ello, existen métricas ya definidos que nos ayudan a obtener dichos resultados, siendo la mayoría de veces de forma objetiva aunque también es útil usar medidas de evaluación subjetiva.

La evaluación de calidad objetiva de la imagen desarrolla medidas cuantitativas que puedan predecir automáticamente la calidad de la imagen percibida. Estas medidas pueden ser usadas para: monitorear dinámicamente y ajustar la calidad de la imagen, optimizar algoritmos y ajustar parámetros de sistemas de procesamiento de imágenes y comparar sistemas de procesamiento de imágenes y algoritmos.

Las métricas de calidad más utilizadas son el Error Cuadrático Medio (MSE, Mean Squared Error) y Relación Señal a Ruido Pico (PSNR, Peak Signal to Noise Ratio). Las medidas objetivas mencionadas anteriormente son altamente atractivos porque son simples de calcular, poco complejos, tienen significados físicos claros, son matemáticamente convenientes en el contexto de la optimización y se consideran como punto de referencia para el desarrollo de otras métricas perceptivas de calidad de imagen/video. Sin embargo, tecnologías de imagen y video digital han demostrado que PSNR presenta algunas limitaciones y que se correlaciona muy pobremente con las medidas subjetivas [Gir96]. Además, se ha demostrado que métricas objetivas de

calidad perceptual como lo es Media de Semejanza Estructural (MSSIM, Mean Structural Similarity) superan al PSNR en la predicción de la calidad de vídeo/imagen subjetiva porque toma en cuenta el sistema visual humano.

En el procesamiento digital de imágenes deben tomarse en cuenta varios aspectos como la percepción psicovisual del ser humano. Éste es un factor importante porque independientemente del tratamiento que se le aplique a una imagen, el observador será quien, según su percepción, decidirá si dicha imagen le agrada o no. Siendo la Puntuación de Opinión Media (MOS, Mean Opinion Score), la mejor métrica para medir la calidad de las imágenes de forma totalmente subjetiva.

A continuación, se realiza una breve descripción de las medidas de calidad de la imagen utilizadas en este trabajo de tesis.

2.3.1. PSNR

PSNR (Relación Señal a Ruido Pico, Peak Signal Noise Ratio) es un término usado en la ingeniería para definir la relación entre la máxima energía posible de una señal y el ruido que afecta a su representación fidedigna. Debido a que muchas señales tienen un gran rango dinámico, PSNR se expresa generalmente en escala logarítmica, utilizando como unidad el decibelio (dB).

El uso más habitual de la medida PSNR es como medida cuantitativa de la calidad de la reconstrucción en el ámbito de los codecs de compresión de imágenes. La señal en este caso son los datos originales, y el ruido es el error introducido por la compresión. Al comparar los codecs de compresión, PSNR es una aproximación a la percepción humana de la calidad de la reconstrucción. Aunque un valor PSNR más alto generalmente indica que la reconstrucción es de mayor calidad [Bal07, GMC10, PS10].

Para definirla se hace indispensable la formulación del MSE, que se calcula al promediar las diferencias de intensidad cuadrada entre los píxeles distorsionados y los píxeles de la imagen de referencia. Dada una imagen libre de ruido monocromática I , de tamaño $M \times N$, y una imagen reconstruida P , la fórmula del MSE se define como

$$MSE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N [I(i, j) - P(i, j)]^2 \quad (2.4)$$

Por lo tanto el PSNR se define como

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \quad (2.5)$$

donde R denota el máximo valor que puede tomar un píxel en la imagen. Cuando el píxel es representado usando 8 bits por muestra, es 255. En ausencia de ruido, las dos imágenes I y P son idénticas, y por lo tanto el MSE es cero. En este caso el valor PSNR es infinito.

2.3.2. MSSIM

MSSIM (Mean Structural Similarity Index Measure), es una medida de similitud promedio entre dos imágenes y se basa en la medida de Índice de Semejanza Estructural (SSIM, Structural, Similarity). SSIM se utiliza para medir la similitud entre dos imágenes. Es una métrica de referencia completa, es decir, la medición o predicción de la calidad de la imagen se basa en una imagen inicial no comprimida o sin distorsión como referencia.

Fue diseñado con el propósito de mejorar los métodos tradicionales, como son PSNR y MSE, que han demostrado ser incompatible con la percepción visual humana. Se diferencia de estas técnicas porque es un modelo basado en la percepción que considera la degradación de la imagen como un cambio percibido en la información estructural, al tiempo que incorpora importantes fenómenos perceptuales, incluyendo el enmascaramiento de la luminancia y los términos de enmascaramiento del contraste. En cambio, PSNR y MSE solo estiman errores absolutos.

La información estructural de una imagen son aquellos atributos que reflejan la estructura de los objetos, siendo independiente de la luminancia media y el contraste de la imagen. Lo que nos lleva a una aproximación de medida de calidad que separe la luminancia, el contraste y las distorsiones estructurales. Los tres componentes son relativamente independientes. Teniendo en conocimiento que x y y son las señales no negativas que han sido alineadas entre sí. Aquí se trabaja con la media y la desviación estándar de una señal ya que son consideradas burdas estimaciones de la luminancia de la señal, mientras que la covarianza (normalizada por la varianza) puede considerarse como una medida de cuanto cambio no lineal posee una señal con respecto a otra.

Comenzamos definiendo la luminancia como

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2.6)$$

donde μ_x es la media de x , μ_y es la media de y y la constante $C_1 = (K_1L)^2$ es incluida para evitar inestabilidad cuando $\mu_x^2 + \mu_y^2$ es muy cercano a cero. Siendo, $K_1 \ll 1$ es una constante, que por lo general resulta ser $K_1 = 0,01$.

La función de comparación de contraste está dado por

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2.7)$$

donde σ_x^2 es la varianza de x , σ_y^2 es la varianza de y y la constante $C_2 = (K_2L)^2$, que como la constante anterior sirve para estabilizar la división, y $K_2 \ll 1$, que por defecto es $K_2 = 0,03$. La variable L , para C_1 y C_2 , es el rango dinámico de los valores de los píxeles que para imágenes de escalas de grises es igual a 255.

La función de comparación estructural es

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (2.8)$$

donde σ_{xy} es la covarianza de x e y . Al igual que las anteriores funciones se introduce una constante $C_3 = C_2/2$.

Finalmente, si combinamos estas 3 ecuaciones (Ecuación 2.6, Ecuación 2.7 y Ecuación 2.8) obtenemos la medida de SSIM entre las señales x y y .

$$SSIM(x, y) = [l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma] \quad (2.9)$$

donde $\alpha > 0$, $\beta > 0$ y $\gamma > 0$ son parámetros usados para ajustar la importancia relativa de los tres componentes. En este caso, $\alpha = \beta = \gamma = 1$. Siendo la Ecuación 2.9 una forma simplificada de

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2.10)$$

Tanto la medida de SSIM como de los tres componentes que la constituyen cumplen las siguientes reglas:

1. Simetría: $SSIM(x, y) = SSIM(y, x)$.
2. Límite: $SSIM(x, y) \leq 1$.
3. Máximo único: $SSIM(x, y) = 1$, si y solo si $x = y$.

SSIM se aplica para evaluar la calidad de imágenes estáticas usando una aproximación de ventana deslizante. El tamaño de la ventana es de 8×8 píxeles. El índice de la medida SSIM se calcula para la región contenida por la ventana que se desplaza pixel a pixel. Lo que resulta en un mapa de índices SSIM de la imagen, que es considerado como el mapa de calidad de la imagen evaluada. Su valor total se define como la media del mapa de calidad, esto es a lo que llamamos índice SSIM medio o MSSIM (Ecuación 2.11) que sirve para evaluar la calidad general de la imagen.

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j) \quad (2.11)$$

donde X es la imagen de referencia e Y la imagen distorsionada; x_j e y_j son las componentes de la imagen en la j -ésima ventana local; y M es el número de ventanas locales en la imagen.

MSSIM demuestra que proporciona una predicción notablemente similar a la de las puntuaciones subjetivas. Sus valores exhiben una consistencia mucho mejor con el aspecto visual cualitativo.

2.3.3. MOS

La ITU (Unión Internacional de Telecomunicaciones) define al puntaje de opinión como el valor en una escala predefinida que un sujeto asigna a su opinión sobre el desempeño de un sistema [ITU06]. Por lo tanto, la Puntuación de Opinión Media (MOS) es una medida de calidad subjetiva y se calcula como el promedio de las puntuaciones de los sujetos. MOS ha surgido como el descriptor más popular de la calidad de los medios percibidos, obteniendo un gran éxito en el dominio de la calidad del habla y también se ha utilizado para otras modalidades como audio, imágenes, video y contenido audiovisual, y en otras numerosas aplicaciones, desde pruebas de laboratorio hasta monitoreo en servicio. Su mayor ventaja es que es reconocido como el método más fiable de evaluación de calidad pero en su contra esta el tiempo, los recursos necesarios, el coste y la imposibilidad de automatizar el proceso.

MOS se expresa como un solo número natural, típicamente en el rango de 1-5, donde 1 es la calidad percibida más baja, y 5 es la calidad percibida más alta. Los niveles de la escala son ordenados por calidad en orden decreciente:

Nivel de calificación	Descripción
5	Excelente
4	Bueno
3	Justo
2	Pobre
1	Malo

Cuadro 2.1: La escala más popular de Calificación de Categoría Absoluta.

Existen otras escalas normalizadas de calificación de la calidad en las recomendaciones del ITU-T [ITU96, ITU08]. En ciertos contextos no hay diferencias estadísticamente significativas entre las calificaciones de los mismos estímulos cuando se obtienen utilizando diferentes escalas.

MOS es calculado como la media aritmética sobre calificaciones individuales realizadas por personas para un estímulo dado en una prueba subjetiva de evaluación de calidad (Ecuación 2.12).

$$MOS = \frac{\sum_{n=1}^N R_n}{N} \quad (2.12)$$

donde R son calificaciones individuales para un estímulo dado por una cantidad de N personas.

Se ha demostrado que en las diferentes escalas de calificación usados por MOS, sus niveles de calificación no son percibidos de forma equidistante por todas las personas. Es decir, que en una escala de Calificación de Categoría Absoluta, una persona puede entender que hay una mayor brecha entre los niveles “Bueno” y “Justo” que entre los niveles “Bueno” y “Excelente”. La distancia percibida entre los niveles de calificación también dependen del idioma en el que se traduce la escala [SWH16]. Además, existe un denominado "sesgo de compensación de escala", en el cual los sujetos a lo largo de un experimento, tienden a dar puntuaciones que abarcan toda la escala de calificación. Esto hace que sea imposible comparar dos pruebas subjetivas diferentes si el rango de calidad presentada difiere. Por lo tanto, las Puntuaciones de Opinión Media obtenidas de diferentes contextos y diseños de prueba no deben compararse directamente.

2.4. Tipografía

Según la Real Academia Española la tipografía es el modo o estilo en que está impreso un texto. La tipografía se clasifica en familias, las mismas que están conformadas por fuentes y estos están unidos por tener características similares [Jua]. Los miembros de una familia, llamadas fuentes tipográficas, se parecen entre sí pero cada uno tiene sus propios rasgos. Las características de cada fuente y familia las hacen únicas y le dan personalidad al texto. Hay que aclarar que existen diversas formas de clasificar a las tipografías, todo porque las letras han ido cambiando debido a la evolución de los medios de comunicación y la tecnología [Che06].

En nuestra investigación solo trabajaremos con tres tipos de tipográficas que son clasificadas según su forma: serifa, sin serifa y máquina de escribir.

2.4.1. Tipografía con serifa

La serifa o remates son pequeños adornos ubicados en los extremos de los trazos de los caracteres. El tipo de letra con serifa o serif se originó según [Cat91] con el alfabeto latino en la antigua Roma, cuando los contornos de las letras eran trazados y cincelados en piedra siguiendo las marcas que gradualmente se hacían más anchas en los extremos y esquinas, creando así las serifas. Además, se caracterizan porque las letras redondas como la o, c, p, b, etc., tienen que ser un poco más grandes ya que visiblemente dan la sensación de ser más pequeñas cuando se agrupan junto con otras letras.

Según los diseñadores, esta tipografía causa tranquilidad, autoridad, dignidad y firmeza. Son adecuados para la lectura de amplios textos ya que los remates y trazos finos ayudan a seguir la lectura más fácilmente, tener una lectura rápida y evitan la

monotonía [Web98]. Algunos ejemplos de fuentes de letra con serifa son: Times New Roman, Georgia, Courier y Garamond.



Figura 2.5: Ejemplo de letra con serifa.

2.4.2. Tipografía sin serifa

Este tipo de tipografía, popularmente conocida como sans-serif, es aquella que no tiene serifas como características de extensión al final de los trazos de los caracteres. Ha sido común en la escritura a lo largo de toda la historia, desde la edad media hasta alcanzar su mayor auge en la edad contemporánea. Entre sus trazos gruesos y delgados no existe contraste, sus vértices son rectos y sus trazos uniformes, ajustados en sus empalmes.

Esta tipografía tiende a reflejar una imagen moderna, limpia y minimalista. En textos impresos las san-serif son usados mayormente en títulos o textos cortos debido a que obligan a esforzar mucho más la vista al leer grandes bloques de texto, dificultando su comprensión [Whe95]. La tipografía sin serifa se ha hecho la más común en medios de comunicación electrónicos, es decir, en televisores, computadores, celulares, etc. debido a que la pixelación en pantallas de baja resolución provocará que los caracteres que contienen detalles finos como las serifas pueden desaparecer o aparentar mayor tamaño [BLM01]. Además, son más adecuados para ser vistos a largas distancias, como es el caso de los anuncios publicitarios.

Algunos ejemplos de fuentes sin serifa son: Arial, Verdana, Tahoma y Helvética.

2.4.3. Tipografía máquina de escribir

Este tipo de tipografía se originó a partir de los caracteres utilizados para escribir en una máquina de impresión personal. Actualmente, la denominación “máquina de escribir” también sirve para definir la apariencia de diversos tipos de letras que nos recuerdan a los que se utilizaron con las máquinas de escribir [Ram16].



Figura 2.6: Ejemplo de letra sin serifa.

Muchos tipos de fuentes nacieron con la máquina de escribir pero sus diseños no convencionales no eran una opción popular y por eso la mayoría desaparecieron con el tiempo junto con la máquina. A pesar de la variedad de estilos creados, la mayoría de los tipos de letra compartían ciertas características. Son conocidos por ser un tipo de fuente monoespaciada, donde sus letras y caracteres ocupan la misma cantidad de espaciado horizontal y tienen el mismo tamaño.



Figura 2.7: Ejemplo de letra del tipo maquina de escribir.

2.5. Métodos de eliminación de ruido

En un proceso de eliminación de ruido se busca restaurar y realzar las imágenes procesadas, ya que el objetivo principal es mejorar la calidad de las imágenes para la percepción humana. La restauración de la imagen es un proceso objetivo que utiliza el conocimiento previo del fenómeno de degradación para recuperar la imagen. Mientras que el realce de la imagen es un proceso subjetivo en el que las opiniones de los individuos están involucradas en el diseño del proceso. A continuación, una breve explicación del proceso de eliminación de ruido [GS13, GW02].

Dada una imagen de entrada $f(x, y)$, el proceso de degradación se modela usando una función de degradación (E) que junto a una operación lineal (aditiva o multiplicativa)

introduce el ruido para producir una imagen ruidosa $g(x, y)$. Ya teniendo $g(x, y)$ con ciertos conocimientos acerca de la función de degradación y el ruido $\eta(x, y)$, el objetivo del proceso de restauración es obtener un estimado $\hat{f}(x, y)$ de la imagen original, es decir, lo que se quiere es que el estimado sea lo más parecido posible a la imagen original de entrada. En general, mientras más conozcamos acerca de E y del ruido η , más parecido será $\hat{f}(x, y)$ a $f(x, y)$. Este proceso se ve reflejado mediante un sencillo esquema mostrado en la Figura 2.8.

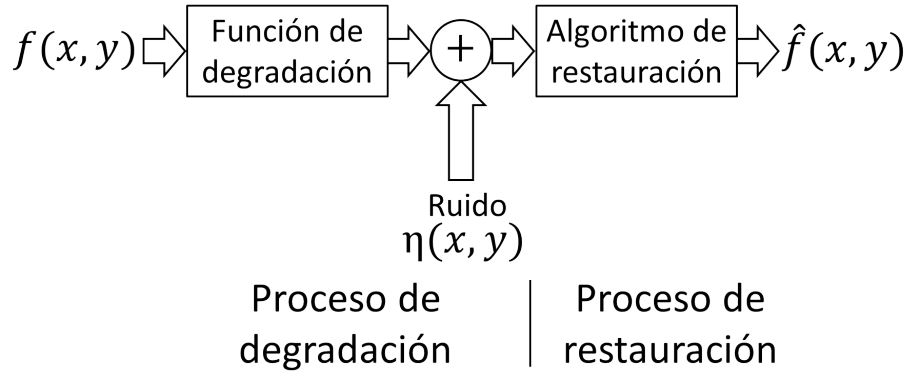


Figura 2.8: Esquema básico de un proceso de eliminación de ruido.

En la actualidad, se están usando transformaciones matemáticas para el proceso de eliminación de ruido en imágenes. Su objetivo es obtener nueva información acerca de la señal que nos brinda el dominio espacial [Bal07, GMC10]. El ejemplo más popular es la Transformada de Fourier que permite realizar un análisis frecuencial de la señal. Sin embargo, una desventaja de Fourier es que solo brinda contenido espectral de la señal y no la localización temporal de los componentes espectrales. Para señales no estacionarias, como son las imágenes, es fundamental realizar el análisis de la señal no solo en dominio frecuencial sino también en el dominio temporal es decir un análisis tiempo-frecuencia.

Es así como van surgiendo la Transformada Wavelet y posteriormente la Transformada Curvelet, las cuales veremos a continuación.

2.5.1. Transformada Wavelet

El término wavelet (onda pequeña) fue introducido en el ambiente matemático por Jean Morlet y Alex Grossmann en 1984 [Can10]. La wavelet es una señal de longitud finita cuyo valor medio es cero y permite el análisis tanto en el tiempo como en la frecuencia alrededor de cierto punto. Otras características son su irregularidad y asimetría a diferencia de una señal sinusoidal. Presentan un buen rendimiento en representar singularidades de punto, debido a que ignoran las propiedades geométricas de las estructuras y no explotan la regularidad de los bordes. Sus principales

características son: transmitir datos comprimidos rápidamente, remover ruido de señales e imágenes, e identificar características cruciales transitorias en conjuntos de datos.

Hay muchas señales que requieren un análisis más flexible, es ahí donde las wavelets pueden representar la señal con otro tipo de funciones que no sean sinusoides, mejorando notablemente la resolución. Además, el análisis es local, permitiendo conocer la respuesta en frecuencia en intervalos de tiempo determinados o regiones de espacio, como en el caso de las imágenes. Todos estos motivos lo convierten en una excelente herramienta para el análisis y procesamiento de señales no estacionarias. En la actualidad hay una gran cantidad de familias de funciones wavelet existentes como la Biortogonal, Coiflets, Symlets, Morlet, etc. pero se recomienda usar, aquella cuya forma se adecue con la señal con la que se está trabajando [Bal07, Can10].

Según [Bal07, Can10, KM02] la Transformada Wavelet consiste en un conjunto de funciones wavelet de duración finita que permiten el análisis de funciones multi-resolución, es decir, descomponen una señal $f(t)$ a través de funciones wavelet que surgen a partir de una función madre. Esta descomposición permite obtener coeficientes susceptibles de interpretación y manipulación. La Transformada Wavelet se define matemáticamente como

$$C(s, \tau) = \int_{-\infty}^{+\infty} f(t) \Psi_{s,\tau}^*(t) dt \quad (2.13)$$

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-\tau}{s}\right) \quad (2.14)$$

donde $\frac{1}{\sqrt{s}} \Psi\left(\frac{t-\tau}{s}\right)$ es la wavelet madre compleja conjugada. Como se observa en la Ecuación 2.14, el resultado está en función de s que representa la escala, es decir, la adaptación de la wavelet a una sección de la señal original y τ es la traslación o desplazamiento de la wavelet. Ya que la Transformada Wavelet permite obtener información relacionada con el tiempo y la frecuencia, su representación gráfica se encuentra sobre un plano llamado tiempo-escala [Mar02], como se muestra en la Figura 2.9.

Con la creación de wavelet madre por Ingrid Daubechies surgen las primeras investigaciones sobre la Transformada Wavelet Discreta, a través de los parámetros s y τ discretos permitiendo el desarrollo de algoritmos más eficientes. Su función está dada por la Ecuación 2.15, que está en función de la wavelet madre discreta que se muestra en la Ecuación 2.16.

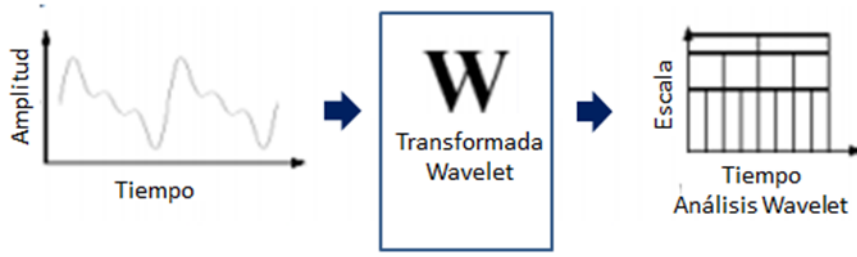


Figura 2.9: Esquema de la Transformada Wavelet.

$$C_{j,k} = \sum_{n \in \mathbb{Z}} f[n] \Psi_{j,k}[n] \quad (2.15)$$

$$\Psi_{j,k}[n] = 2^{-\frac{j}{2}} \Psi[2^{-j}n - k] \quad (2.16)$$

Es este tipo de transformada con la que se trabaja debido a su facilidad de implementación en los computadores. Es capaz de entregar suficiente información tanto para el análisis como para la reconstrucción de una señal con una significativa reducción de tiempo de procesamiento y es mucho más fácil de implementar con respecto a su versión continua.

2.5.1.1. Transformada Wavelet en 2D

La Transformada Wavelet es usada para el procesamiento digital de imágenes y con su versión discreta se logra reducir la gran cantidad de espacio que requieren las imágenes de gran resolución. En el caso de imágenes, la Transformada Wavelet Discreta se utiliza dos veces, es decir, una transformada para las filas y otra para las columnas. Este proceso de operaciones de dilatación y adaptación es equivalente a aplicar un banco de filtros sobre la imagen.

El proceso de encontrar los coeficientes wavelet y obtener una representación tiempo-escala se llama análisis multi-resolución [Bal07, Can10, Mar02]. En este proceso se filtra una señal en el dominio del tiempo mediante filtros pasa bajos y pasa altos con distintas frecuencias de corte para analizar la señal en diferentes escalas. Esta operación de descomposición se repite para las señales resultantes del filtrado anterior. El filtrado permite cambiar la resolución de la imagen mientras que la escala se cambia mediante operaciones de sub-muestreo e interpolación. El sub-muestreo consiste en reducir la tasa de muestreo o eliminar muestras de la señal, que no afecta a la resolución y no se pierde información, mientras que la interpolación incrementa la tasa de muestreo agregando nuevas muestras a la señal. Luego de haber

descompuesto la señal en niveles predefinidos anteriormente, obtenemos un grupo de señales que representan la misma señal pero en diferentes bandas de frecuencias.

Al momento de juntarlas, debido a que el principio de incertidumbre no nos permite conocer que frecuencias existen en un tiempo dado, solo podemos conocer que bandas de frecuencias existen en un determinado intervalo de tiempo. Es por ello que la Transformada Wavelet permite el uso de intervalos grandes de tiempo en aquellos segmentos en los que se requiere mayor resolución en baja frecuencia y regiones más pequeñas (mayor resolución en el tiempo) donde se requiere información en alta frecuencia.

Este análisis multi-resolución crea una descomposición piramidal de la imagen en varias escalas y una descomposición de los coeficientes en varios niveles (descomposición multi-resolución) que se logra con filtros pasa bajos y pasa altas. Al aplicar la Transformada Wavelet Discreta en una imagen, en el primer nivel se obtienen cuatro coeficientes: aproximaciones, detalles horizontales, detalles verticales y detalles diagonales. La aproximación (frecuencias bajas) es la que contiene mayor cantidad de información, mientras que los detalles de orientación (frecuencias altas) son valores cercanos a cero debido al filtrado que se realiza sobre la imagen original.

En la Figura 2.10, muestra la distribución gráfica de la pirámide de descomposición según los coeficientes resultantes, el primer nivel (V1, D1, H1 donde V, D, H representan las direcciones vertical, diagonal y horizontal, respectivamente) representa la banda de más alta frecuencia y el nivel más fino de resolución. En cambio el último nivel de descomposición (en este caso nivel tres) corresponde a la banda de frecuencia más baja y de resolución más burda. Al recorrer los niveles de baja resolución a alta resolución, se observa una disminución de energía en las capas. En la Figura 2.11, se muestra un ejemplo real de la pirámide de descomposición.

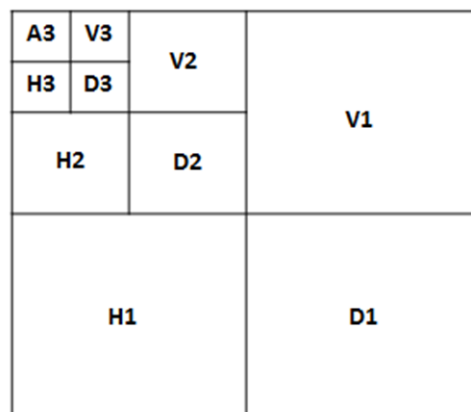


Figura 2.10: Descomposición piramidal de coeficientes wavelet a tres niveles [Bal07].

A todo este proceso de descomponer las señales en detalles y aproximaciones se le conoce como “análisis”. Mientras que el proceso de reconstrucción o mejor conocido



Figura 2.11: a) Imagen original de Barbara (izquierda), b) Descomposicion a dos niveles (derecha) [Bal07].

como “síntesis”, es aquel que a partir de los detalles y aproximaciones se encarga de reconstruir la señal. Este proceso es llevado a cabo por la Transformada Wavelet Discreta Inversa, como se observa en la Figura 2.12 donde el sub-muestreo está representado por círculos con flechas hacia abajo y la interpolación está representado por círculos con flechas hacia arriba. Llevan un número 2 en su interior ya que el banco de filtros basados en el algoritmo Mallat cuenta con la desventaja de aumentar al doble el número de datos originales por lo que el costo matemático y computacional también incrementa.

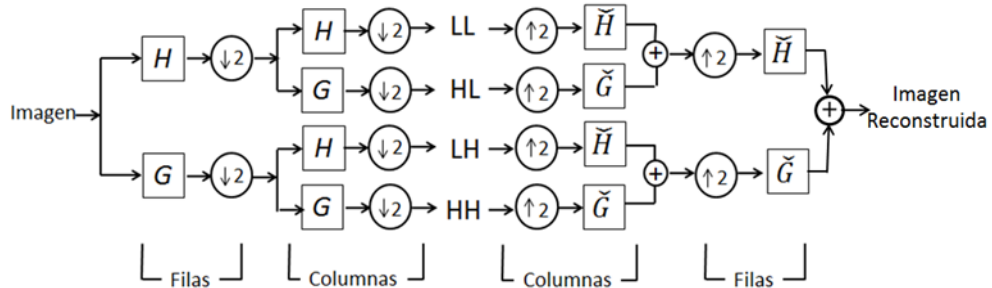


Figura 2.12: Análisis y síntesis en 2D de un solo nivel [Can10].

2.5.2. Transformada Curvelet

En los últimos años se han desarrollado nuevas herramientas matemáticas y computacionales basadas en ideas multi-escala y multi-resolución, como las wavelets, pero se fue demostrando que esta idea clásica de multi-resolución está lejos de ser efectiva. A pesar de que las wavelets son buenas tratando con singularidades de punto, estas

no son buenas en obtener representaciones direccionales que capturen líneas anisotrópicas y bordes, mientras proporcionan descomposiciones poco densas. Hay que tener en cuenta que en el procesamiento de señales, los fenómenos más interesantes ocurren en curvas, es decir, los bordes de las imágenes.

Es ahí donde surge una nueva familia de transformada geométrica multi-escala llamada curvelet que fueron introducidas por Candes y Donoho [CD00] para encontrar una representación eficiente a los objetos con discontinuidades a lo largo de sus curvas, resolver el problema de compresión de datos en imágenes y sobre todo intentan superar las limitaciones de representaciones multi-escala tradicionales que tiene la wavelet.

Según [GMC10, PVL12, MP10], curvelet es una transformada direccional multi-escala que permite una representación dispersa no adaptativa de objetos con bordes, es decir, es una técnica no adaptativa para la representación de objetos multi-escala. Esta transformada equivale a llevar a cabo al mismo tiempo dos análisis: análisis multi-resolución y un análisis geométrico.

Al trabajar en un plano de dos dimensiones en tiempo continuo, la curvelet es localizada por una variable en el dominio espacial x (la posición) y una variable en el dominio de la frecuencia ω (la escala). Esta última maneja coordenadas polares r y θ . Se empieza con un par de ventanas llamadas ventana radial $W(r)$ y ventana angular $V(t)$. Ambos son suaves, no negativos y de valor real. W toma argumentos positivos reales donde $r \in (1/2, 2)$, y V toma argumentos reales, donde $t \in [-1, 1]$.

A continuación, se introduce una ventana de frecuencia U_j definida en el dominio de Fourier como se muestra en la Ecuación 2.17, donde $[j/2]$ es la parte entera de $j/2$. Por lo tanto, el soporte de U_j es una cuña polar (es decir, una ventana cartesiana en coordenadas polares) que esta definida por el soporte de la ventana radial y ventana angular.

$$U_j(r, \theta) = 2^{-3j/4} W(2^{-j}r) V\left(\frac{2[j/2]\theta}{2\pi}\right) \quad (2.17)$$

Por otra parte, la familia de curvelets se construye mediante la aplicación de dilataciones, rotaciones y traslaciones de la función curvelet madre $\varphi_j(x)$ a escala 2^{-j} , donde $j = 0, 1, 2, \dots$. Teniendo en cuenta esto, se introduce [CDDY05, CD02]:

- La secuencia equidistante de ángulos de rotación definida como $\theta_\ell = 2\pi \times 2^{-[j/2]} \cdot \ell$, siendo $\ell=0,1,2, \dots$ tal que $0 \leq \theta_\ell \leq 2\pi$. Note que el espaciado entre ángulos consecutivos es dependiente de la escala.
- La secuencia de parámetros de traslación definida como $k = (k_1, k_2) \in \mathbb{Z}^2$.

Debido a lo anterior podemos definir a las curvelet (Ecuación 2.18) como funciones de $x = (x_1, x_2)$, a escala 2^{-j} , orientación θ_ℓ y posición $x_k^{(j,\ell)} = R_{\theta_\ell}^{-1}(k_1 \times 2^{-j}, k_2 \times 2^{-j/2})$

$$\varphi_{j,\ell,k}(x) = \varphi_j(R_{\theta_\ell}(x - x_k^{(j,\ell)})) \quad (2.18)$$

donde R_θ significa la matriz de rotación en θ radianes y R_θ^{-1} es su inversa. Las posiciones $x_k^{(j,\ell)}$ están en diferentes cuadrículas regulares para cada diferente ángulo de rotación y estas cuadrículas tienen un espaciado diferente en las dos direcciones siendo consistente con la escala parabólica, es decir, con la relación de ángulos y escalas. Entonces un coeficiente curvelet, que está definido como Ecuación 2.19, es el producto interno entre una función f y una curvelet $\varphi_{j,\ell,k}$.

$$c(j, \ell, k) = \int_{R^2} f(x) \overline{\varphi_{j,\ell,k}(x)} dx \quad (2.19)$$

Los elementos de la familia curvelet ahora deberían proveer un embañosado en el plano de frecuencia 2D basada en coordenadas polares. Por consiguiente, la construcción de las curvelets considera coordenadas polares en el dominio de la frecuencia y construyen elementos curvelet apoyándose localmente cerca de cuñas (Figura 2.13), donde el número de cuñas es $N_j = 4 \times 2^{\lfloor j/2 \rfloor}$ a escala 2^{-j} , es decir, se duplica en cada segundo anillo circular.

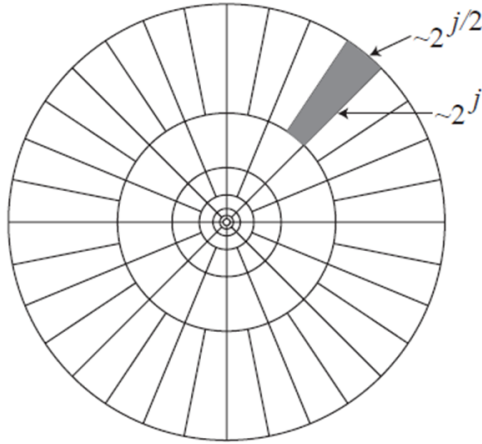


Figura 2.13: Embaldosado del dominio de la frecuencia en cuñas para la construcción de curvelets y el área sombreada representa una cuña (wedge) genérica [CDDY05].

2.5.2.1. Transformada Curvelet Discreta

Según [CDDY05, YDC], la Transformada Curvelet Discreta toma como entrada un array cartesiano (imagen en 2D) de la forma $f[m, n]$, donde $0 \leq m \leq M$ y $0 \leq n \leq N$, y nos permite representar su salida como un conjunto de coeficientes $c^D(j, \ell, k)$, definido como

$$c^D(j, \ell, k) = \sum_{0 \leq m \leq M} \sum_{0 \leq n \leq N} f[m, n] \varphi_{j, \ell, k}^D[m, n] \quad (2.20)$$

donde $\varphi_{j, \ell, k}^D$ (D de digital) es una forma de onda digital que preserva las propiedades de la curvelet continua, donde j es la escala, ℓ es la orientación y k los parámetros de locación espacial. Esta aproximación curvelet implementa la ley de escala parabólica en las sub-bandas en el dominio de la frecuencia para capturar con mayor eficacia aquellos bordes curvos de una imagen.

En la definicion de tiempo continuo de la ventana $U_{j, \ell}$, esta extrae suavemente las frecuencias cerca de la corona diádica $\{2^j \leq r \leq 2^{j+1}\}$ y cerca del ángulo $\{-\pi 2^{-j/2} \leq \theta \leq \pi 2^{-j/2}\}$. Pero en la Transformada Curvelet Discreta es conveniente reemplazar estos conceptos por equivalentes cartesianos; aquí la “corona cartesiana” estará basada en trapezoides en vez de círculos como se observa en la Figura 2.14.

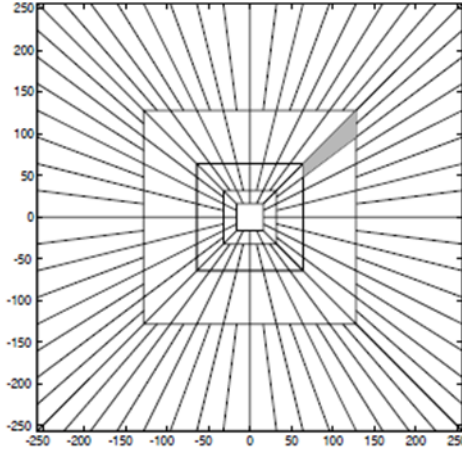


Figura 2.14: Plano seccionado de la Transformada Curvelet con soporte pseudo-polar parabólico en el plano de frecuencia [CDDY05].

Lamentablemente, debido a que esta transformada usa una compleja serie de pasos que involucran el análisis de Ridgelet en la Transformada de Radon, generaba

que su rendimiento sea extremadamente lento. Por tal motivo se descartó usar la Transformada de Radon, con lo que se logró reducir la redundancia e incrementar la velocidad en la transformada. Con el propósito de hacer más fáciles de usar y comprender las curvelets se rediseñaron proponiendo nuevos métodos de Transformada Curvelet.

2.5.2.2. Transformada Curvelet Discreta Rápida

En [CDDY05, SMEG12, SKVR12] se hablan de nuevas formas de Transformada Curvelet basadas en diferentes operaciones de muestras de Fourier que pueden ser implementadas de dos formas: vía USFFT (Transformada de Fourier Rápida Desigualmente Espaciada, Unequally-Spaced Fast Fourier Transform) y vía wrapping. En USFFT los coeficientes curvelet se encuentran mediante el muestreo irregular de los coeficientes de Fourier de una imagen. Mientras que con la transformada via wrapping aplica una serie de traslaciones y una técnica envolvente basado en la envoltura (wrapping) de muestras de Fourier especialmente seleccionadas. Ambos retornan una tabla de coeficientes curvelet digitales indexados por un parámetro de escala, orientación y locación espacial. Su ventaja con otras transformaciones digitales es que son más sencillas, rápidas y menos redundantes ya que son versiones mejoradas de la Transformada Curvelet, recibiendo el nombre de Transformadas Curvelet Discretas Rápidas.

La principal diferencia entre las nuevas transformadas es la cuadrícula espacial que utilizan para trasladar las curvelets en cada escala y ángulo [CDDY05]. En la versión wrapping, la rejilla es la misma para todos los ángulos dentro de cada cuadrante, sin embargo, a cada curvelet se le da la orientación apropiada. Como resultado, la transformación via wrapping es más rápida, robusta, simple de entender y aplicar. La transformada basada en wrapping es una pirámide multi-escala (no-estándar) que consiste en diferentes orientaciones y posiciones a un nivel de baja frecuencia, que permite una representación poco densa, no adaptativa y casi óptima de objetos con singularidades a lo largo de sus suaves curvas [CDDY05]

Según [SMEG12, SKVR12], lo que hacen estas nuevas transformadas en el dominio espectral es utilizar las ventajas de la Transformada Rápida de Fourier, en el cual tanto la curvelet como la imagen a cierta escala y orientación son transformadas en el dominio de Fourier. Al final de este proceso se obtiene un conjunto de coeficientes curvelet al aplicar la Transformada Rápida de Fourier Inversa (IFFT, Inverse Fast Fourier Transform) sobre el producto entre el FFT de la curvelet y el FFT de la imagen [CD99]. Este grupo se compone de coeficientes curvelet en orden ascendente de las escalas y orientaciones. El proceso de extracción de características completo utilizando un único curvelet ilustrado en Figura 2.15a, donde la imagen puede ser recuperada sin error desde los coeficientes generados invirtiendo los pasos involucrados.

Sin embargo, hay un problema en el espectro de frecuencia obtenido al aplicar la

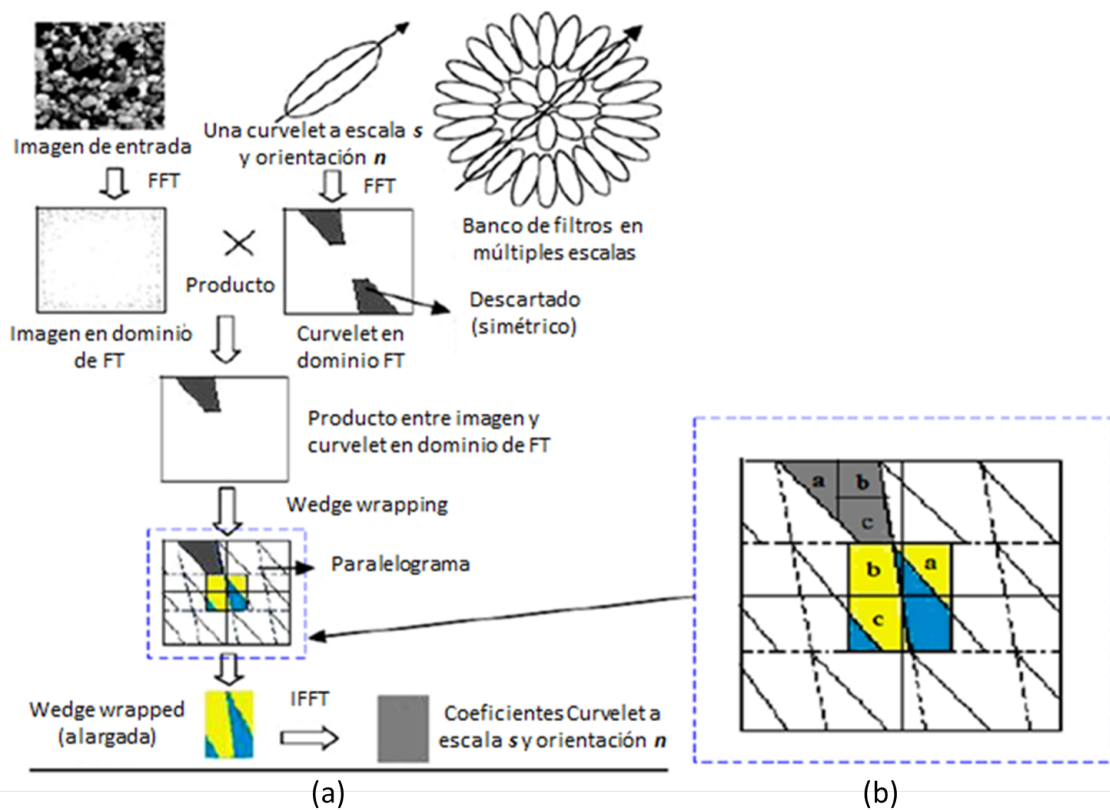


Figura 2.15: Diagrama de proceso de FDCT [SMEG12].

IFFT. La respuesta en frecuencia de la curvelet es una cuña trapezoidal que necesita ser envuelta (wrapped) en un soporte rectangular para que se pueda realizar la IFFT. La envoltura de la cuña trapezoidal se logra embaldosando periódicamente el espectro dentro de la cuña y luego juntando el área de coeficiente rectangular en el origen. A través de este plano seccionado periódico, la región rectangular recolecta las porciones fragmentadas correspondientes de la cuña de los paralelogramos de los alrededores. A este proceso se le conoce como la transformada vía wrapping, que se puede observar en la Figura 2.15b.

Como se observa en la Figura 2.15b, a fin de lograr la IFFT en la Transformada de Fourier de la cuña, la cuña tiene que ser convertida en un rectángulo. La idea principal es replicar la cuña en una cuadrícula 2D, y el rectángulo en el centro logra capturar todos los componentes a , b , y c de la cuña. El wrapping de cuñas se realiza en todas las cuñas a cada escala en el dominio de la frecuencia, con lo que se obtiene un conjunto de sub-bandas o cuñas en cada nivel de descomposición curvelet. Estas sub-bandas son la colección de coeficientes de curvelet discreta.

2.6. Factorización de Matrices no-Negativa

El reciente crecimiento de flujo de información y complejidad de datos pueden volverse inmanejables, por lo que es necesario algoritmos adecuados de pre-procesamiento para reducir las dimensiones de los datos y así poder extraer, visualizar y codificar la información deseada. Su objetivo es obtener representaciones de baja dimensionalidad con una mínima pérdida de información descubriendo características más importantes de los datos. Actualmente, existen muchos procedimientos para la reducción de dimensionalidad que utilizan la factorización de matrices.

2.6.1. Definición

NNF es una técnica de análisis de datos multivariado lineal que busca estimar componentes físicamente significativos latentes basados en valores no negativos, donde no se permite la substracción. Según [Hoy04], si consideramos que nuestros datos consisten en M mediciones de N variables escalares no negativas. En donde, los vectores de medida (N dimensiones) están representados como v^m ($m = 1, \dots, M$) podemos formular una representación lineal aproximada de los datos mostrada en la Ecuación 2.21

$$v^m \approx \sum_{i=1}^K w_i h_i^m = W h^m \quad (2.21)$$

donde W es una matriz $N \times K$ que contiene los vectores base w_i como sus columnas. Hay que tener en cuenta que cada vector de medición se escribe en términos de los mismos vectores base. Los K vectores base pueden considerarse como los "bloques constructivos" de los datos, y el vector de coeficientes h^m (de M dimensiones) describe la fuerza con que cada bloque de construcción está presente en el vector de medida v^m . Así es como se describe la definición lineal de NMF.

2.6.2. Modelo Básico

Según [Hoy04, CZPA09, LS01, CZ06], NMF es un algoritmo de aprendizaje no supervisado que descompone una matriz de observaciones V de dimensiones $N \times M$ donde cada elemento $v_{ij} \geq 0$. Aquí la matriz original tiene N filas almacenando las observaciones y M columnas que representan las dimensiones. Se puede reconstruir dicha matriz original de forma aproximada por medio del producto entre las matrices W (matriz base desconocida) y H (matriz de componentes ocultos desconocidos) de dimensiones $N \times K$ y $K \times M$, respectivamente. Donde sus elementos $w_{ij} \geq 0$ y $h_{ij} \geq 0$ son positivos, tal que el algoritmo NMF pueda encontrar matrices no negativas y el rango de factorización reducido de $K \ll \min(N, M)$, donde K es el

número de vectores base y es requerido como entrada para la mayoría de los algoritmos NMF. Su determinación es algo complejo y depende de la matriz de datos y de la aplicación. Mientras mayor sea el valor de K , mejor es la aproximación de los datos pero un valor pequeño de K nos lleva a un modelo menos complejo. Al tener un valor de componentes K nos aseguramos que W y H sean menores que la matriz original V tal que

$$V \approx W \times H \quad (2.22)$$

Escrito en esta forma, se hace evidente que una representación de datos lineales es simplemente una factorización de la matriz de datos.

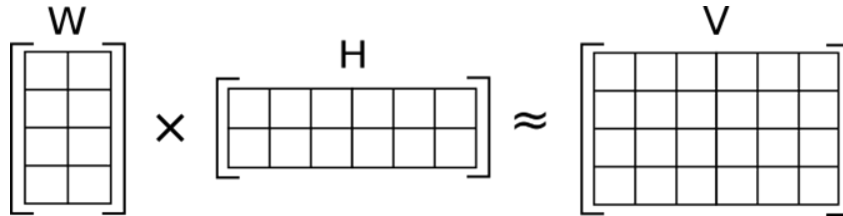


Figura 2.16: La multiplicación entre W y H nos dan una reconstrucción aproximada de V .

Al hacer que W y H sean más pequeños que V , se vuelven más fáciles de almacenar y manipular. Otra razón para factorizar V en matrices más pequeñas, es que se puede representar aproximadamente los elementos de V con mucho menos datos. Una representación gráfica de NMF se muestra en Figura 2.16.

De esta forma es que NMF logra descomponer y obtener una aproximación de una matriz original a través del producto de dos matrices inferiores, es decir, es una representación lineal de datos aproximados con la propiedad que todas las matrices no tienen elementos negativos [Hoy04, SZCS12]. NMF puede ser considerada como una representación basada en partes, donde un valor de cero representa la ausencia y un número positivo representa la presencia de un evento o componente.

La importancia de la restricción de la no negatividad se debe a que la mayoría de los datos en la vida real son no negativos y sus componentes fundamentales ocultos solo tiene un significado físico cuando estos son no negativos; además, permite una representación puramente aditiva y hace que las matrices resultantes sean más fáciles de inspeccionar. Por ejemplo, la intensidad de los píxeles en una imagen. Por otro lado, la restricción de la no negatividad estimula a una representación dispersa de los datos, ya que según [CZPA09] es importante cuando se quiere que tales datos sean representados con la menor cantidad de componentes relevantes.

2.6.3. Funciones de coste

Para conseguir la factorización de matrices se logra mediante una minimización iterativa de una determinada función de coste, que cuantifican el error de aproximación entre V y WH , [SZCS12, LG15, Sch10], como lo son: la distancia euclídea (Ecuación 2.23) o la divergencia Kullback-Leibler (KL) (Ecuación 2.24)

La distancia euclídea cuenta con la condición de restricción $\|W\|_2 = 1$, y que W y H no sean negativas, está definida como:

$$D_{EU}(V \parallel WH) = \sum_{ij} (V_{ij} - (WH)_{ij})^2 \quad (2.23)$$

Mientras que la divergencia Kullback-Leibler sirve para medir la similitud entre la matriz V y el producto de $W \times H$, las restricciones son las mismas que la anterior y su función es:

$$D_{KL}(V \parallel WH) = \sum_{ij} \left[V_{ij} \log \frac{V_{ij}}{(WH)_{ij}} - (V_{ij} - WH_{ij}) \right] \quad (2.24)$$

A pesar de que son las más populares no son las únicas opciones que existen. Hay diferentes funciones de coste que asignan diferentes penalizaciones o restricciones. Entre las que podemos mencionar: las divergencias de la familia Bregman [DS05], medida de información de Renyi [Dev05], divergencias Csiszar [CZA06a], divergencia Kompass [Kom07], la divergencia $-\alpha$ [CLKC08], la divergencia Itakura-Saito [FBD09] entre otras.

2.6.4. Reglas de aprendizaje multiplicativo

Por otro lado, para lograr la factorización a través de la minimización de la función de coste D dada por la Ecuación 2.23 y Ecuación 2.24 con respecto a $W \geq 0$ y $H \geq 0$, no se puede calcular en forma cerrada, sino que se tiene que encontrar mediante un proceso iterativo alternando entre actualizar una matriz mientras se mantiene la otra fija.

En [LS99, LS01] se proponen las siguientes reglas de actualización multiplicativa para la función de coste cuadrático D_{EU}

$$H_{kj} \leftarrow H_{kj} \frac{[W^T V]_{kj}}{[W^T W H]_{kj}} \quad W_{ik} \leftarrow W_{ik} \frac{[X H^T]_{ik}}{[W H H^T]_{ik}} \quad (2.25)$$

y reglas para la divergencia KL generalizada D_{KL}

$$H_{kj} \leftarrow H_{kj} \frac{\sum_i W_{ik} V_{ij} / [WH]_{ij}}{\sum_l W_{lk}} \quad W_{ik} \leftarrow W_{ik} \frac{\sum_j H_{kj} V_{ij} / [WH]_{ij}}{\sum_p H_{kp}} \quad (2.26)$$

y se probó que estas actualizaciones multiplicativas nunca aumenta la función de coste respectiva.

Otro factor importante es la inicialización del algoritmo iterativo a partir del cual se obtienen las matrices W y H . Al ser NMF una técnica basada en gradientes sufren de quedarse atascados en mínimos locales por lo que se recomienda ejecutar el algoritmo muchas veces usando diferentes y aleatorias inicializaciones. Es por ello que la solución y convergencia de NMF depende mucho de la inicialización por lo que se debe elegir matrices iniciales W y H adecuadas sino la eficiencia del algoritmo se verá afectada. Una mala inicialización nos llevará a un mínimo local que nos conducirá a una mala solución. Hay que tener en cuenta que una buena inicialización para un conjunto de datos no es necesariamente buena para otros conjuntos de datos.

Las propiedades de convergencia dependen del algoritmo en particular. En general, la convergencia global es muy difícil de probar, pero se puede demostrar que la mayoría de los algoritmos no aumentan la función de coste en cada iteración. Usualmente, la convergencia se prueba midiendo la disminución de la función de coste entre iteraciones sucesivas, y el algoritmo se detiene si la disminución cae por debajo de cierto umbral predeterminado.

2.6.5. Restricciones adicionales

El concepto de codificación dispersa significa que sólo unas pocas unidades de una gran población se utilizan realmente para representar vectores de datos típicos. Esto implica que la mayoría de las unidades toman valores pequeños y sólo unos pocos adoptan valores significativamente no nulos al representar un elemento de datos.

Por ejemplo, Li [LHZC03] usa representaciones espacialmente localizadas de las imágenes en la función de coste local NMF (LNMF, Local Non-Negative Factorization)

$$D_{lnmf} = \sum_i \sum_j (V_{ij} \ln \frac{V_{ij}}{[WH]_{ij}} - V_{ij} + [WH]_{ij}) + \alpha \sum_{ia} [W^T W]_{ia} - \beta \sum_k [HH^T]_{kk} \quad (2.27)$$

En este caso, a la divergencia KL se le añade un término que penaliza las entradas grandes en WW^T así como un término que penaliza pequeñas entradas diagonales de HH^T . El término W debería incrementar la dispersión de los pesos y alentar soluciones donde sólo unos pocos componentes de base son necesarios para representar

cada observación, mientras que la última expresión debería hacer que los vectores de base sean lo más ortogonales posible a fin de minimizar la redundancia entre diferentes vectores base.

En resumen, lo que se necesita para realizar un modelo básico de NMF es:

1. Función de coste. Ante todo se tiene que definir una función de coste que permita medir la discrepancia entre las matrices V y WH .
2. Algoritmo de optimización. Una vez elegida la función de coste que incluye las restricciones adicionales opcionales, existen varias estrategias numéricas posibles para resolver el problema de optimización.

2.6.6. Aplicaciones

La decomposición de un conjunto de datos a gran escala de cantidades no negativas juega un papel importante en muchas áreas. Desde su invención por Paatero y Trappér a mediados de los noventa bajo el nombre Factorización de Matrices Positivas [PT94], decenas de investigadores aplicaron NMF a una gran variedad de aplicaciones en las diferentes áreas de investigación. Más tarde, Lee y Seung establecieron el ahora conocido NMF en 1999 [LS99], fue ahí donde nació la gran popularidad de este método en diferentes campos.

Ahora nombraremos algunas aplicaciones en las cuales se ha usado las diferentes variaciones de NMF.

2.6.6.1. Minería de Textos

Según [CZPA09, Wit] la minería de textos, que involucra la clasificación de documentos de texto en grupos o clusters acorde a su similitud en características semánticas, trata de buscar patrones en el texto. Por ejemplo, un motor de búsqueda web por lo general retorna miles de páginas en respuesta a una consulta amplia, haciendo difícil a los usuarios de encontrar o identificar información relevante. Es ahí donde los métodos de clustering son usados para agrupar los documentos encontrados en una lista de temas significativos, siendo uno de ellos la Indexación Semántica Latente que implica la Descomposición de Valor Singular que no garantiza que todos los componentes de los vectores singulares sean no negativos.

Es ahí donde aparece el NMF en lo que respecta a la minería de textos ya que es usado para extraer temas o características semánticas de colecciones de documentos y exhibe mejor discriminación para el clustering de los datos que se superponen parcialmente. En [CZPA09] indica que NMF es superior a Indexación Semántica Latente en términos de precisión y que su rendimiento es comparable a métodos jerárquicos.

2.6.6.2. Vigilancia de correo electrónico

NMF también ha sido exitosamente aplicado a la extracción y detección de los temas de los mensajes de los correos electrónicos. La idea de fondo de agrupar (clustering) características semánticas de mensajes de correos electrónicos es similar al visto anteriormente con el clustering de documentos.

En [BB05] hacen uso del NMF para clasificar mensajes de correo electrónico de una empresa. Su objetivo es identificar los temas importantes de los mensajes sin la necesidad de leer dichos mensajes y clasificarlos de acuerdo a las características semánticas.

2.6.6.3. Aplicaciones biomédicas

NMF también ha encontrado muchas aplicaciones relevantes en el procesamiento y análisis de datos biomédicos. Muchos de ellos para la clasificación de la expresión del gen, como por ejemplo el análisis de datos de micro-matrices de ADN [PTGM04, GC05], donde los datos no negativos corresponden a los niveles de expresión genética y NMF fue aplicado exitosamente para la clasificación del cáncer. Otras aplicaciones biomédicas incluyen separación y clasificación de señales electroencefalograma [LZL04, CCR06], reconocimiento de pliegues de proteínas [OP06] y disciplinas de formación de imágenes como la resonancia magnética [SDB⁺04, DMSS07], tomografía de emisión de positrones [LLCL01] y espectroscopia de fluorescencia [GPH04].

2.6.6.4. Clasificación de imágenes

La clasificación de imágenes es una parte esencial de un patrón u objeto de un sistema de reconocimiento y usualmente está basado en un aprendizaje supervisado o no supervisado. Muchas aplicaciones se han enfocado en el NMF aplicado sobre el reconocimiento de rostros [BP04] y problemas de clasificación de imágenes (reconocimiento de dígitos manuscritos).

Por ejemplo, en el reconocimiento de rostros, los vectores característicos son aprendidos de un conjunto de caras frontales distinguidos por diversas características como expresiones faciales, condiciones de iluminación, etc. Un número selecto de esas caras forman un conjunto de entrenamiento. Las imágenes de entrenamiento son usualmente escaneadas en orden lexicográfico para formar los vectores columna en V , luego V es descompuesta con NMF en una matriz base W y una matriz de pesos H . Debido a sus propiedades, muchas veces NMF supera a métodos basados en el Análisis de Componente Principal (PCA, Principal Component Analysis) para el reconocimiento de rostros debido a que las imágenes obtenidas con NMF pueden ser consideradas como representaciones basadas en partes poco densas y locales como lo son los ojos, cejas, orejas y boca [LS01].

2.6.6.5. Otras aplicaciones

Otras menciones importantes son la clasificación y reconocimiento de sonidos [SB03, SS05], donde las características acústicas son extraídas de grabaciones de sonidos, por ejemplo, los patrones de instrumentos específicos y resolver el problema de separación de fuente acústica. Datos financieros [DRdFC07], para identificar las tendencias subyacentes en los datos bursátiles. En las áreas de ciencias ambientales y quimiometría, donde se analizaron espectros o conjuntos de concentraciones [PT94] y aplicaciones meteorológicas [ST09]. NMF también ha sido usado en el área de la criptografía [XYF08] o en el análisis de un proceso de grabado [RMRM08].

3 Estado del Arte

Las imágenes digitales son propensas a ser contaminadas por diferentes tipos de ruidos durante los procesos de adquisición y transmisión; provocando una degradación en la calidad de la imagen. Existen muchas técnicas para remover el ruido en las imágenes y por lo general son técnicas basadas en filtros lineales en el dominio espacial y la frecuencia. Aunque estas logran eliminar el ruido en imágenes digitales, muchas de las imágenes no logran mantener los bordes y difuminan la imagen, haciendo que la eliminación de ruido en imágenes que contengan texto sea aún más difícil. Es por ello que se vuelven populares aquellas técnicas no lineales y que se basan en el dominio de la transformación: Wavelet y Curvelet; o simplemente provocan el nacimiento de otras propuestas, como aquellos basados en NMF.

Para poder tener una idea clara de cómo implementar y desarrollar los métodos que hemos trabajado en esta investigación, nos hemos guiado de ciertos trabajos previos que se enfocan en la tarea de la eliminación de ruido en imágenes. A continuación, se nombrará y hablará de forma concisa de aquellos trabajos que ayudaron a cumplir con el desarrollo de nuestra investigación.

En el trabajo [PS10] se propone una técnica multi-resolución con umbralización suave basado en la estimación de varianza local para la eliminación de ruido en imágenes. Con ayuda de una técnica de umbral adaptativo logran reducir el ruido en la imagen preservando los bordes. Su método es implementado usando la Transformada Curvelet Discreta del tipo wrapping ya que es más sencilla, rápida y menos redundante; así logran superar las limitaciones que tiene la Transformada Wavelet. Usan el proceso de umbralización suave porque produce imágenes visualmente más agradables en comparación a la umbralización dura. Para calcular el umbral, utilizan la aproximación BayesShrink, donde el valor del umbral adaptativo es determinado para diferentes sub-bandas dependiendo de la relación entre la varianza de ruido y la desviación estándar de la señal.

Por otro lado, el valor de la varianza de ruido ($\hat{\sigma}^2$) la estiman a través de un estimador de mediana robusto, mostrado en la Ecuación 3.1 (Transformada Wavelet) y Ecuación 3.2 (Transformada Curvelet).

$$\hat{\sigma} = \frac{\text{Mediana}(|g_{i,j}|)}{0.6745} \quad g_{i,j} \in \text{sub-banda} \quad (3.1)$$

$$\hat{\sigma} = \frac{\text{Mediana}(|g\{j\}\{l\}|)}{0.6745} \quad g\{j\}\{l\} \in \text{sub} - \text{banda} \quad (3.2)$$

El proceso de eliminación de ruido es evaluada usando medidas tales como MSE y PSNR. Concluyen que el algoritmo BayesShrink de umbral suave preserva mejor los bordes y mejora el suavizado de la imagen. Además, confirman que la Transformada Curvelet se desempeña mejor que la wavelet, ya que presenta una representación dispersa, lo que permite rastrear mejor las curvas y por lo tanto elimina mejor el ruido que wavelet.

En el trabajo realizado por [PVL12] se enfoca sobre lo problemático que puede llegar a hacer el ruido para el reconocimiento de caracteres en aplicaciones OCR y lo importante que es eliminarlo previamente. Presentan, implementan y comparan tres métodos que eliminan ruido. La imagen contaminada con ruido gaussiano e impulsivo es binarizada.

El primer método basado en wavelet, comienza encontrando la Transformada Wavelet de 2 niveles en la imagen contaminada de ruido. Utilizan la wavelet Haar ya que no difumina la imagen y calculan σ (Ecuación 3.3) usando la Desviación Absoluta Mediana (MAD, Median Absolute Deviation) de los coeficientes wavelet de alta frecuencia (coeficientes de detalle) en su primer nivel de descomposición

$$\sigma = \frac{MAD}{0.6745} = \frac{\text{mediana}(|\omega_k|)}{0.6745} \quad (3.3)$$

donde ω_k representa los coeficientes wavelet en la primera escala. Luego calculan el umbral λ (para cada nivel) basado en el umbral universal, que está dado por

$$\lambda_j = \sigma \sqrt{2 \log(N_j)} \quad (3.4)$$

donde N_j es el tamaño de la matriz de coeficientes wavelet en su j -ésimo nivel. Aplican la umbralización dura a los coeficientes wavelet para cada nivel. Finalmente, aplican la Transformada Wavelet Inversa para obtener la imagen sin ruido y usan el método Otsu para binarizarla.

El segundo método está basado en el método shrinkage de wavelet mostrado en [dSMS12]. Tras aplicar la Transformada Wavelet, sobre la imagen se encuentra un mapa de bordes (una imagen binaria) a partir de los coeficientes de aproximación aplicando detección de bordes basado en el método Canny. Utilizan estos coeficientes de baja frecuencia debido a que la mayor parte del ruido se encuentran en los detalles (coeficientes de altas frecuencias). Así determinan los coeficientes wavelet con borde

y sin borde. La umbralización para los coeficientes con borde es casi tres veces mayor que la umbralización de los coeficientes sin borde, de este modo se logra preservar los bordes de los textos.

El tercer método es propuesto con el propósito de poder conservar mejor los bordes de los caracteres incluso en presencia de alto ruido. Primero, encuentran los coeficientes curvelet de la imagen contaminada con ruido y calculan la escala. Obtiene σ y λ para cada escala. La desviación estándar del ruido, también la calculan usando la Ecuación 3.3; sin embargo, el umbral λ_j es calculado como:

$$\lambda_j = K\sigma\sqrt{2\log(N_j)} \quad (3.5)$$

donde, N_j es el tamaño de la matriz de coeficientes curvelet en su j -ésima escala y K es un parámetro ajustable para encontrar un mejor PSNR. Después, la umbralización dura y la Transformada Curvelet Inversa para obtener la imagen eliminada de ruido. Aplican el método Otsu global y luego realizan una operación morfológica estándar como post-procesamiento para remover los pixeles espurios para suavizar los bordes de los caracteres y obtener un fondo limpio y claro.

Los tres métodos son puestos a prueba con varios tipos de imágenes de documentos con diferentes resoluciones, fuentes y tamaños de letra, que son contaminados con diferentes niveles de ruido gaussiano, impulsivo y una mezcla de ambos. Los resultados muestran que ambos métodos basados en wavelet presentan un buen rendimiento con bajos niveles de ruido pero a medida que el ruido aumenta el rendimiento del método basado en curvelet saca ventaja. El método basado en curvelet presenta un buen rendimiento a un nivel de ruido extremo $\sigma = 100$ y muestra buenos resultados al eliminar una mezcla de ruido gaussiano e impulsivo, lo que lo hace más versátil cuando se trata de eliminar ruido. Además, comprueban que el software OCR es capaz de reconocer la mayoría de caracteres con niveles de ruido menores a $\sigma = 75$. Concluyen que usar curvelets para la eliminación de ruido concede la ventaja de preservar las características del borde de los caracteres en las imágenes de documentos contaminados con ruido lo que lo hace conveniente para las aplicaciones OCR.

En [GMC10], para evitar la degradación de antiguos manuscritos y escrituras históricas hindús, combaten el ruido usando un método basado en curvelet y operadores morfológicos de apertura y clausura para eliminar el fondo del documento mediante la preservación de regiones de primer plano o regiones de fondo, respectivamente. Mencionan que las curvelets ofrecen dos características importantes: la anisotropía y direccionalidad. Por otro lado, aunque las wavelets se usan para separar las singularidades puntuales, las curvelets durante la extracción de características bidimensionales son capaces de describir discontinuidades de la imagen a lo largo de un borde (curvas y líneas) con un número mínimo de coeficientes [CD99]. En resumen estos son los pasos más importantes del método propuesto:

- Aplicar la Transformada Curvelet para calcular los coeficientes curvelets
- Aplicar la umbralización
- Aplicar la Transformada Curvelet Inversa
- Aplicar los operadores morfológicos de apertura y cierre

Para la experimentación usaron la toolbox de curvelet (CurveLab-2.1.3) implementado en MATLAB [CDDY15] e hicieron uso de una cámara de alta resolución para digitalizar los documentos históricos. Utilizan PSNR para medir el rendimiento y calidad del método propuesto entre la imagen original y la imagen mejorada con los operadores morfológicos. Usan alrededor de 100 imágenes de prueba (manuscritos hindús). Tras la experimentación se concluye que al aplicar los operadores morfológicos de apertura y clausura, los niveles PSNR en las imágenes de documentos históricos mejoran bastante en comparación a los documentos que solo fueron tratados con curvelet.

Por último, en [SZCS12] se usa un novedoso método de eliminación de ruido en imágenes usando Factorización de Matriz No-Negativa Local (LNMF) añadiéndole una condición de restricción dispersa, llamada SC-LNMF (NMF Local con Restricción Dispersa, Sparse Constraint LNMF), que puede extraer características espacialmente localizadas de forma eficiente.

En esta investigación nos explican que las partes positivas aprendidas no están necesariamente localizadas usando NMF, entonces para asegurar la localización de las características de las imágenes se utiliza el algoritmo LNMF. A comparación del NMF, LNMF añade tres restricciones (máxima dispersidad, máxima expresividad y máxima ortogonalidad) en la función de coste y tiene como meta aprender características localizadas que son adecuados para tareas de reconocimiento de imágenes. Además, las bases LNMF son más adecuadas para una representación localizada que NMF. Todo esto lo hace eficiente al momento de extraer características locales y eliminar el ruido de las imágenes.

Lamentablemente, LNMF no tiene en cuenta la distribución dispersa de la imagen ni tampoco el control de dispersidad de los vectores base de características sobre los coeficientes de dispersidad. Es por ello que para mejorar la dispersidad de la matriz característica y la localidad de las características de la imagen proponen el algoritmo SC-LNMF. Además cuenta con cierta condición de restricción de dispersidad, que permite extraer características localizadas de forma eficiente. De este modo, se logra eliminar el ruido utilizando los vectores base de características SC-LNMF y la regla de umbral de shrinkage basada en [Sha08, Hyv97].

El algoritmo LNMF con restricción de dispersidad considera: El error de reconstrucción de la imagen, la densidad de los coeficientes de característica, la máxima expresividad de la matriz de pesos W y la máxima ortogonalidad de la matriz H . Durante su proceso hacen uso del algoritmo de descenso por gradiente para actualizar W y H en turnos. Primero, se mantiene fija la matriz de características base

W y así actualizan la matriz de coeficientes H . Posteriormente, se fija la matriz H y luego se actualiza W .

Durante la experimentación ponen a prueba SC-LNMF usando 10 imágenes de la base de datos de huella palmar, reduciendo su tamaño a 128×128 píxeles para facilitar los cálculos. Los tres algoritmos basados en NMF previamente son entrenados y para evaluar la calidad de la eliminación de ruido emplean la medida SNR.

Las observaciones más resaltantes durante la extracción de características son que cuanto mayor son las dimensiones de la matriz característica W habrá una mejor localización de los vectores base de características SC-LNMF. Además, también se observa que la restricción de dispersidad y la localidad de los vectores base de características de SC-LNMF eran apenas mejores que las de NMF y LNMF. Además, sin importar el tipo de algoritmo se encontró que mientras mayor son las dimensiones de la matriz característica W , más lenta era la velocidad de convergencia. En cuanto a lo que se refiere a la eliminación de ruido, el método SC-LNMF es más eficiente en comparación que NMF y LNMF (que usan el método shrinkage).

4 Metodología de Solución y Diseño

Como ya sabemos, el ruido deteriora la calidad de las imágenes durante los procesos de adquisición y transmisión. Por lo que el objetivo principal de los diferentes métodos de eliminación de ruido en imágenes es remover el ruido sin distorsionar la imagen procesada, es decir, obtener el mejor estimado posible de la imagen de entrada. Al inicio, los métodos más usados fueron aquellos que se basan en el filtrado lineal. Con el pasar del tiempo fueron surgiendo muchos métodos no-lineales para cumplir y mejorar dicho objetivo.

A continuación, se mostrará el proceso de tres diferentes métodos de eliminación de ruido en imágenes de documentos con los cuales se está trabajando en esta investigación. Siendo el primero un esquema basado en Wavelet, el segundo método basado en Curvelet y el tercer método, propuesto en este trabajo, basado en NMF.

4.1. Método basado en wavelet

Uno de los primeros métodos usados para la eliminación de ruido en imágenes fue wavelet debido a sus ventajas de localización tanto en el dominio de la frecuencia como en el dominio espacial, su análisis multi-resolución y análisis multi-escala [Bal07, PS10, CDDY05]. Las wavelets funcionan bien sólo al representar las singularidades puntuales, ya que ignoran las propiedades geométricas de las estructuras y no explotan la regularidad de los bordes. Así que nos servirá como punto de partida para poder comparar sus resultados con los otros métodos.

La imagen de entrada $f(x, y)$ pasará por los siguientes procedimientos:

- a) Se añade el tipo de ruido que se quiere eliminar: gaussiano, impulsivo y multiplicativo.
- b) La señal ruidosa $g(x, y)$ de tamaño $M \times N$, se descompone en dos niveles usando la Transformada Wavelet.
- c) Se estima el umbral universal (λ) en cada nivel de descomposición usando la Ecuación 4.1, que será utilizada el proceso de umbralización dura [DJ94] a los coeficientes de la imagen ruidosa.
- d) Se aplica la Transformada Wavelet Inversa en dos niveles para obtener la imagen sin ruido $\hat{f}(x, y)$
- e) Finalmente, se aplica el proceso de binarización $\hat{f}_{bin}(x, y)$ usando el método Otsu [Ots79].

Se debe tener en cuenta que durante el proceso de la Transformada Wavelet se usa como función madre a la wavelet Haar, debido a que no produce difuminación sobre las imágenes. En este caso prevalece la umbralización dura sobre la umbralización suave debido a que la umbralización suave tiende a remover algunos coeficientes de borde creando difuminación en los bordes de texto, perdiendo así detalles finos. Lamentablemente, la umbralización dura presenta la desventaja que en niveles altos de ruido genera caracteres con bordes de texto irregulares. Hay que tener en cuenta que la umbralización dura iguala a cero los coeficientes del vector cuya magnitud sean menor (en valor absoluto) que el umbral calculado.

Por otro lado, el umbral universal [Don95] se calcula usando la Ecuación 4.1, donde N_j es el tamaño de la matriz de coeficientes wavelet en su j -ésimo nivel.

$$\lambda = \sigma \sqrt{2 \log(N_j)} \quad (4.1)$$

El umbral esta en función de σ que es la desviación estándar de ruido dada por la Ecuación 4.2, donde MAD es la Desviación Mediana Absoluta de los coeficientes wavelet de alta frecuencia ω_k a escala uno, es decir, en el primer nivel de descomposición.

$$\sigma = \frac{MAD}{0,6745} = \frac{\text{mediana}(|\omega_k|)}{0,6745} \quad (4.2)$$

Se hace uso del método Otsu, que es un algoritmo que convierte una imagen de escala de grises a binaria. Otsu asume que la imagen contiene dos clases de píxeles (píxeles de primer plano y píxeles de fondo), siguiendo un histograma bimodal. Su objetivo es encontrar el umbral óptimo para la segmentación de la imagen, bajo el criterio de máxima varianza entre el plano de fondo y el objeto [Ots79]. El método Otsu es muy utilizado en reconocimiento de patrones, binarización de documentos y visión por computadora.

El esquema del método basado en wavelet se muestra en la Figura 4.1

4.2. Método basado en curvelet

Debido a que los coeficientes wavelet no capturan demasiada información direccional y fallan al dar una representación dispersa a lo largo de las curvas, en [CD00] introducen la Transformada Curvelet que sirve para representar los bordes y otras singularidades a lo largo de curvas mucho más eficientemente que las transformaciones tradicionales.

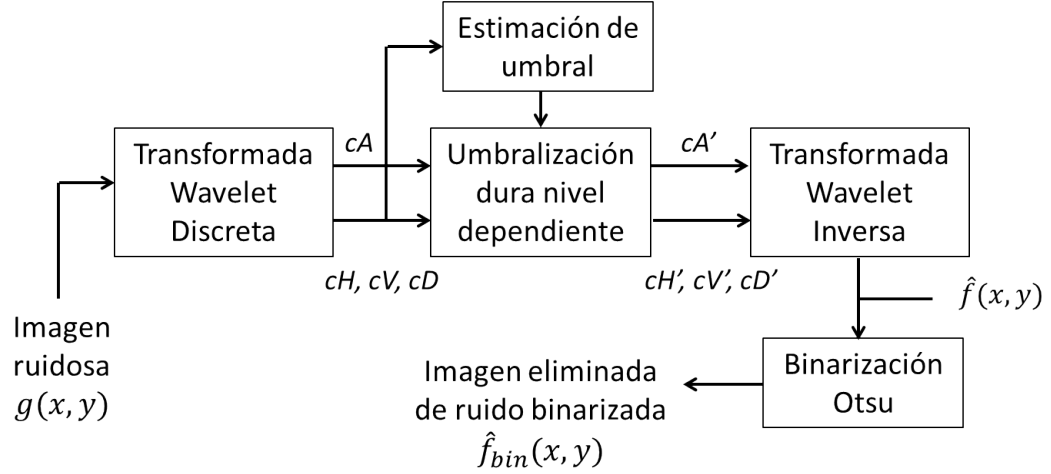


Figura 4.1: Diagrama de bloques para el proceso de eliminación de ruido en imágenes basado en Transformada Wavelet [PVL12].

Se diseñó el siguiente esquema basados en [PVL12, SCD02, YDM10], donde se utilizó la umbralización adaptativa basados en curvelet que con eficiencia se estima y elimina los ruidos de las imágenes digitales. A continuación, se detallarán los principales pasos:

- a) Se añade el tipo de ruido que se quiere eliminar: gaussiano, impulsivo y multiplicativo.
 - b) Se aplica la Transformada Curvelet Discreta Rápida (FDCT) vía wrapping sobre el ruido $\eta(x, y)$.
 - c) Se calcula la desviación estándar σ para cada coeficiente curvelet ruidoso usando la Ecuación 4.4.
 - d) Se aplica la Transformada Curvelet Discreta Rápida (FDCT) vía wrapping sobre la imagen ruidosa $g(x, y)$.
 - e) La umbralización dura [DJ94] se realiza sobre los coeficientes curvelet de la imagen ruidosa.
 - f) Se aplica la Transformada Curvelet Discreta Inversa para obtener la imagen eliminada de ruido $\hat{f}(x, y)$.
 - g) Se binariza la imagen eliminada de ruido $\hat{f}_{bin}(x, y)$ usando el método Otsu [Ots79].
- Si la imagen de entrada tiene dimensiones $M \times N$, entonces el número de escalas se calcula usando la Ecuación 4.3.

$$L = \text{ceil}(\log_2(\min(M, N)) - 3) \quad (4.3)$$

donde *ceil* se refiere al comando de MATLAB que redondea el número al entero mayor más próximo. Habiendo calculado los coeficientes curvelet ruidosos, ahora se requiere calcular la desviación estándar σ_{jl} de cada uno, mostrada en la Ecuación 4.4. Para ello se hará uso del valor MAD donde los coeficientes dependen de una escala (j) y ángulo (l). Los coeficientes curvelet y desviaciones estándar ruidosos nos servirán para establecer el umbral durante el proceso de umbralización dura. El umbral será $3\sigma_{jl}$ para todos los niveles menos para la escala más fina donde es $4\sigma_{jl}$ [AM14].

$$\sigma_{jl} = \frac{\text{Mediana}(|g\{j\}\{l\}|)}{0,6745} \quad g\{j\}\{l\} \in \text{sub-banda} \quad (4.4)$$

El esquema del método basado en curvelet se muestra en la Figura 4.2.

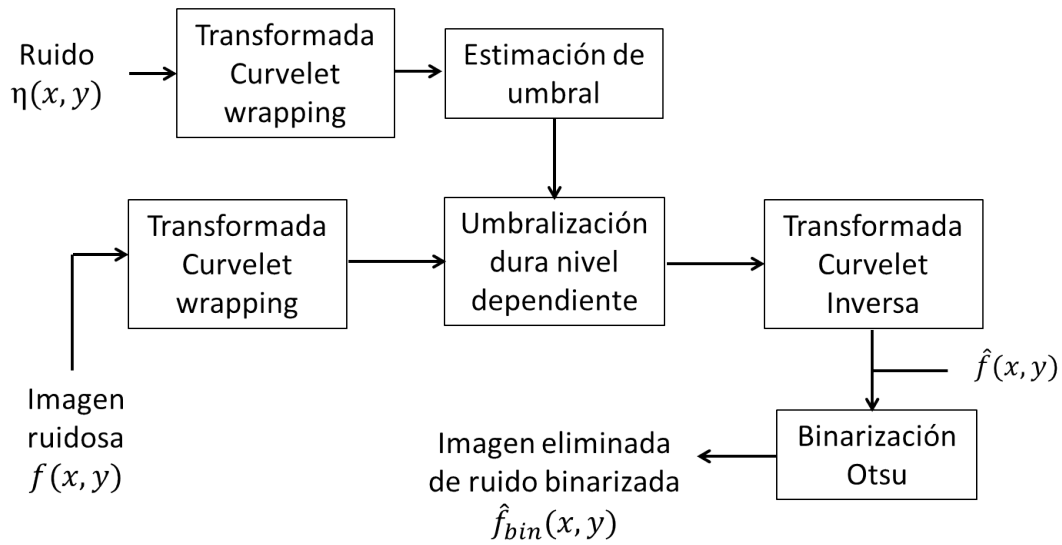


Figura 4.2: Diagrama de bloques del proceso de eliminación de ruido en imágenes usando el método basado en Transformada Curvelet.

4.3. Método basado en NMF

En esta sección se realizará la eliminación de ruido a través del algoritmo NMF, tomando como referencia el trabajo realizado en [LG15] donde se aplica NMF para la eliminación de ruido en voz mediante la descomposición de la señal en una combinación de vectores base usando la minimización de una función de coste, como es la divergencia Kullback-Leiber.

Usaremos la divergencia KL como nuestra función de coste primordial ya que presenta una serie de ventajas como: mejores resultados perceptuales, buenas propiedades

de convergencia, es adecuada cuando los datos presenta un amplio rango dinámico [BDSh] y buen coste computacional [SLWE09]. Además, permite un control explícito del grado de dispersión en las descomposiciones NMF.

Para encontrar el valor óptimo local entre V y WH usando la divergencia Kullback-Leiber, se procedió a usar un esquema iterativo con reglas de aprendizaje multiplicativo [LS99], establecido como

$$W \leftarrow W \otimes \frac{\frac{V}{WH} H^T}{1 H^T} \quad H \leftarrow H \otimes \frac{W^T \frac{V}{WH}}{W^T 1} \quad (4.5)$$

donde 1 es una matriz, donde todos sus elementos son unos, de tamaño V , y las multiplicaciones \otimes y divisiones son componente a componente.

NMF no asume dispersión de datos entre las columnas de W pero suele proporcionar una descomposición dispersa [LS99] de los datos logrando así una interpretabilidad más fácil. Para conseguir un control sobre esa representación dispersa, nos basamos en [CZPA09, CZA06b] para la función de coste KL, donde se logra regularizar al algoritmo NMF utilizando proyecciones no lineales sobre la Ecuación 4.5. Tras aplicar este proceso, las reglas de aprendizaje obtenidas quedan establecidos de la siguiente manera

$$W \leftarrow \left[W \otimes \frac{\left[\frac{V}{WH} H^T \right]^\omega}{1 H^T} \right]^{(1+\alpha_\omega)} \quad H \leftarrow \left[H \otimes \frac{\left[W^T \frac{V}{WH} \right]^\omega}{W^T 1} \right]^{(1+\alpha_h)} \quad (4.6)$$

donde $\alpha_\omega \geq 0$ y $\alpha_h \geq 0$ son parámetros de regularización o también conocidos como factores de dispersidad y $\omega \in (0, 2)$ es un parámetro de relajación que también controla el grado de dispersión y además la velocidad de convergencia del algoritmo. Al tener en cuenta los parámetros de regularización, el exponente de las reglas de aprendizaje será mayor que uno, provocando que los valores más pequeños sean aún más cercanos a cero y los valores grandes se incrementen a medida que el número de iteraciones aumenta [LG12].

Hay que tener en cuenta que el proceso de inicialización es muy importante debido a que la función de coste puede tener muchos mínimos locales. De acuerdo a [CZPA09] las funciones de costes de la Ecuación 2.23 y Ecuación 2.24 son estrictamente convexas con respecto a una sola de las matrices W o H pero nunca a ambas. Para resolver este problema, las matrices W y H se inicializaron usando el algoritmo de inicialización múltiple [CZPA09].

Ya teniendo estas consideraciones en cuenta, NMF se aplicará a la imagen ruidosa representado como V_{mix} , como la combinación lineal de los componentes que representan la imagen limpia W_{clean} y el ruido W_{noise} , mejor conocidos como vectores

base. La Figura 4.3 muestra una representación simple de la imagen ruidosa (V_{mix}), donde el vector base de la imagen W_{clean} y sus coeficientes de activación H_{clean} se usan para reconstruir la imagen limpia $V_{clean} \approx W_{clean} H_{clean}$, mientras que los vectores base del ruido W_{noise} y sus coeficientes de activación H_{noise} reconstruyen la señal de ruido $V_{noise} \approx W_{noise} H_{noise}$.

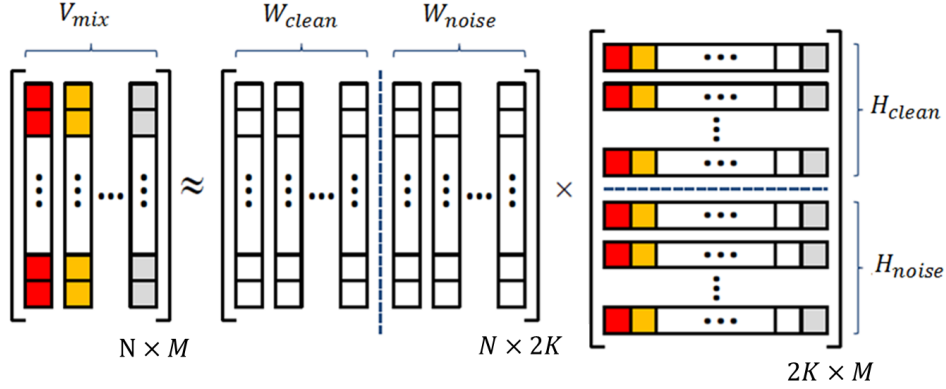


Figura 4.3: Representación básica sobre una imagen ruidosa al ser descompuesta por NMF [LG15].

Todo el proceso se divide en dos etapas: entrenamiento y eliminación de ruido [LG15].

Previamente, es muy importante efectuar un apropiado proceso de inicialización de las matrices factorizadas (W y H) como ya se mencionó anteriormente. Aquí se usará el algoritmo de inicialización dado en [CZPA09], para ello se generarán 50 pares de matrices aleatorias uniformes (W_0 y H_0) y se escogerá la factorización más pequeña que produce la distancia euclídea entre V y $(W_0 H_0)$. Luego, las matrices iniciales se refinarán a través de la minimización de la divergencia KL entre V y $(W_0 H_0)$ usando las reglas de aprendizaje multiplicativo de la Ecuación 4.5, donde el punto de parada del algoritmo será el número máximo de iteraciones, que en nuestra experimentación resultó ser 1000.

4.3.1. Entrenamiento

Al aplicar NMF sobre los datos de entrenamiento de la imagen limpia (V_{clean}) y el ruido (V_{noise}), se obtendrán sus correspondientes vectores base (W_{clean} y W_{noise}). Para ello, primero se minimiza la divergencia KL entre V_{clean} y V_{noise} y sus respectivas matrices factorizadas $W_{clean} H_{noise}$ y $W_{noise} H_{noise}$, respectivamente, usando las reglas de aprendizaje multiplicativo dadas en la Ecuación 4.5. Así es como se obtienen los vectores base de la imagen W_{clean} y el ruido W_{noise} , que serán usados en el proceso de eliminación de ruido como modelos de imagen y ruido.

En nuestra experimentación se tomó el 60 % de las imágenes de la base de datos de

cada resolución para el modelado de las imágenes, mientras que para el ruido se construyó la misma cantidad de imágenes (ruido gaussiano, impulsivo y multiplicativo según sea el caso) cuya varianza varía en el rango de 0.01 a 0.1.

4.3.2. Eliminación de Ruido

Luego de la obtención de nuestros vectores base W_{clean} y W_{noise} estos no necesitan ser reentrenados, por lo que se mantienen fijos y son concatenados para formar un solo vector base llamado W_{all} , construyendo de este modo el modelo de la imagen ruidosa ya que contiene componentes de imagen y ruido. Dada la imagen ruidosa V_{mix} , se calcula su factorización $W_{all}H_{all}$ minimizando la divergencia KL entre V_{mix} y $W_{all}H_{all}$ (donde $W_{all} = [W_{clean} \ W_{noise}]$ y $H_{all} = [H_{clean} \ H_{noise}]$). Puesto que la matriz W_{all} permanece fija en esta etapa, las reglas de aprendizaje con control de restricción de dispersidad (Ecuación 4.6) solo se aplican sobre la matriz H_{all} con los parámetros (ω y α_h) adecuados. Donde en [LG15, SZCS12] NMF usa la divergencia KL con una serie de factores que controlan la dispersión de las matrices factorizadas, debido a que una representación dispersa puede mejorar el proceso de eliminación de ruido. Una vez recalculado la matriz H_{all} , la estimación de la imagen limpia resulta ser $V_{clean} \approx W_{clean} H_{clean}$, siendo H_{clean} las filas de H_{all} correspondientes a los coeficientes de W_{clean} .

Para conocer el impacto de los diferentes parámetros NMF sobre la calidad de la imagen se realizaron experimentos previos. Los parámetros considerados fueron el número de vectores base y los parámetros de regulación ω y α_h . Todo esto, usando la función de coste basado en la divergencia KL ya que se obtiene mejores resultados en términos de PSNR. A continuación, se presentarán los resultados obtenidos de este análisis previo:

- El número de vectores se estableció entre un rango de 10 a 150 en intervalos de 10. Donde los resultados de PSNR y MSSIM son muy bajos cuando el número de vectores base varía entre 10 a 70. En cambio, cuando se prueba con 80 o más vectores base los valores de PSNR y MSSIM aumentan, siendo los resultados muy similares. En nuestro caso se eligió 100 vectores base debido a que con este valor se aprecia un ligero aumento en los resultados de PSNR y MSSIM.
- En cuanto a los valores que controlan la dispersión de la matriz de activaciones, se alternó ω desde 1 hasta 2.5 y α_h desde 0 hasta 1.2 [LG15]. Donde los mejores resultados PSNR y MSSIM se obtuvieron con la combinación $\omega = 1,75$ y $\alpha_h = 0$.

Después de las etapas de entrenamiento y eliminación de ruido basado en NMF, se utiliza una máscara espectral que nos conduce a una mejor estimación para cada fuente de la señal mixta, es decir, obtendremos un mejor resultado de la imagen que ya ha sido procesada por NMF. Para ello, utilizamos estimaciones de V_{clean} y V_{noise} para construir una máscara de la siguiente manera:

$$M = \frac{\hat{V}_{clean}^p}{\hat{V}_{clean}^p + \hat{V}_{noise}^p} \quad (4.7)$$

donde $p > 0$ y la división son operaciones elemento a elemento. Nótese que, los elementos de $M \in (0, 1)$, y el uso de diferentes valores p conduce a diferentes tipos de máscaras. Estas máscaras escalarán cada componente de la señal mixta según la frecuencia con una relación que explica cuánto de cada fuente contribuye en la señal mixta tal que:

$$\hat{V}_{clean} = M \otimes V_{mix} \quad (4.8)$$

$$\hat{V}_{noise} = (1 - M) \otimes V_{mix} \quad (4.9)$$

donde \hat{V}_{clean} y \hat{V}_{noise} son las matrices finales estimadas que representan la imagen limpia y el ruido, 1 es una matriz de unos y \otimes es multiplicación elemento a elemento. Usando esta idea nos vamos a asegurar que la suma de las dos matrices estimadas sea igual a la señal mixta [GE11]. Finalmente la imagen limpia estimada es binarizada \hat{V}_{clean} .

En la Figura 4.4 se presenta un esquema resumen de todo el proceso explicado anteriormente para la eliminación de ruido basado en NMF.

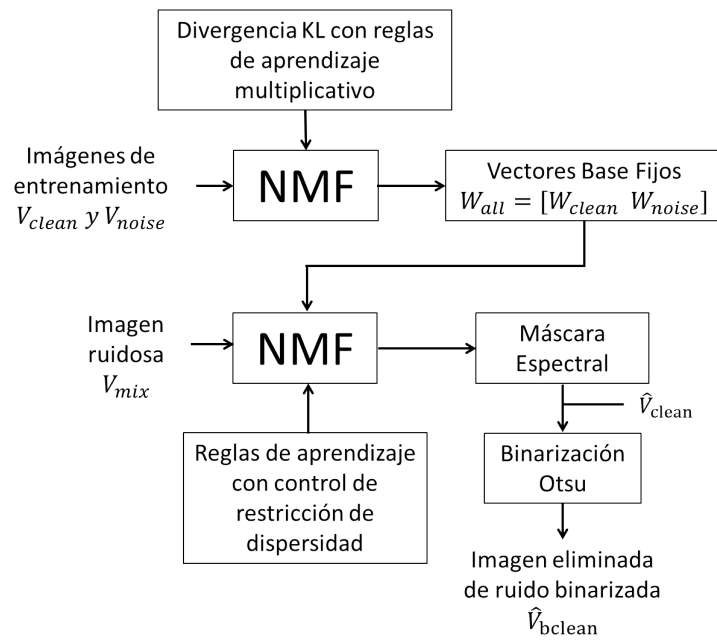


Figura 4.4: Diagrama de bloques del proceso de eliminación de ruido usando el método basado en NMF con control de restricción de dispersidad.

5 Experimentación y Análisis

La base de datos usada para la experimentación consiste en imágenes creadas por R.M.J. Castro-Bleda, S. España-Boquera, J. Pastor-Pellicer y F. Zamora-Martínez, brindadas por la página web de competencias Kaggle, disponible en [Inc17]. La base de datos consta de 147 imágenes (conformada por 51 imágenes con resolución de 258×530 píxeles y 96 imágenes de 420×530 píxeles) habiendo diferenciado en esta base de datos 6 fuentes tipográficas, como se puede observar en Figura 5.1, pero para esta investigación se clasificaron en 3 familias tipográficas: serifa, sin serifa y máquina de escribir. Nuestro análisis mostrará cómo actúan los diferentes métodos de eliminación de ruido en imágenes de documentos usando tres familias tipográficas con diferentes tipos y niveles de ruido.

Con la finalidad de comprobar el rendimiento de los diferentes métodos de eliminación de ruido, las imágenes de documentos fueron contaminadas artificialmente con ruido digital, tales como: ruido gaussiano, impulsivo y multiplicativo de media cero y varianza entre 0.01 a 0.1.

En nuestra investigación mediremos el rendimiento de los tres métodos de eliminación de ruido en imágenes de documentos mediante tres medidas de evaluación de calidad: PSNR, MSSIM y MOS. Siendo esta última una medida totalmente subjetiva ya que son personas las que evalúan la calidad de las imágenes.

Asimismo, se calculó las diversas medidas de evaluación de calidad para las imágenes contaminadas con los diferentes tipos y niveles de ruido digital considerados, con el fin de corroborar y observar la funcionalidad de los métodos implementados. A estos experimentos le vamos a llamar "Línea base".

Nuestra experimentación se divide en 2 etapas:

- Primero, realizamos las mediciones usando PSNR y MSSIM.
- Segundo, realizamos las mediciones usando MOS.

5.1. Mediciones basadas en PSNR y MSSIM

PSNR mide la relación entre la máxima energía posible de una señal y el ruido que afecta a su representación original, y se usa como medida cuantitativa de la calidad de reconstrucción de imágenes [GMC10, Bal07, PS10]. Para lo cual se encuentra primero el valor MSE que representa el error cuadrado acumulativo entre la imagen

There exist several methods to design to be filled in. For instance, fields may bounding boxes, by light rectangles or b These methods specify where to write and mize the effect of skew and overlapping w the form. These guides can be located on a paper that is located below the form or th directly on the form. The use of guides on is much better from the point of view of scanned image, but requires giving more more importantly, restricts its use to tasks

(a)

There exist several methods to design fo be filled in. For instance, fields may be surr ing boxes, by light rectangles or by guiding ru ods specify where to write and, therefore, n of skew and overlapping with other parts o guides can be located on a separate sheet located below the form or they can be print form. The use of guides on a separate she from the point of view of the quality of th but requires giving more instructions and, restricts its use to tasks where this type of a

(c)

There exist several methods to c fields to be filled in. For insta be surrounded by bounding boxes, b or by guiding rulers. These metho to write and, therefore, minimize and overlapping with other parts o guides can be located on a separat that is located below the form or directly on the form. The use of sheet is much better from the poin quality of the scanned image, but

(e)

There exist several methods to design to be filled in. For instance, fields may be bounding boxes, by light rectangles or b These methods specify where to write and mize the effect of skew and overlapping w the form. These guides can be located on a paper that is located below the form or th directly on the form. The use of guides on is much better from the point of view of scanned image, but requires giving more more importantly, restricts its use to tasks

(b)

There exist several methods to design fo be filled in. For instance, fields may be surr ing boxes, by light rectangles or by guiding ru ods specify where to write and, therefore, n of skew and overlapping with other parts o guides can be located on a separate sheet located below the form or they can be print form. The use of guides on a separate she from the point of view of the quality of th but requires giving more instructions and, restricts its use to tasks where this type of a

(d)

There exist several methods to d fields to be filled in. For instar be surrounded by bounding boxes, by or by guiding rulers. These method to write and, therefore, minimize t and overlapping with other parts of guides can be located on a separat that is located below the form or t directly on the form. The use of g sheet is much better from the point quality of the scanned image, but r

(f)

Figura 5.1: Seis diferentes tipografías de la base de datos: a) Letra con serifa itálica, b) Letra con serifa, c) Letra sin serifa itálica, d) Letra sin serifa, e) Letra máquina de escribir itálica y f) Letra máquina de escribir.

Tipografía	Cantidad de imágenes de resolución $258 \times 530px$ / $420 \times 530px$	TOTAL
Con Serifa	29 / 57	86
Sin Serifa	31 / 70	101
Máquina de Escribir	35 / 63	98

Cuadro 5.1: Cantidad de imágenes a prueba por tipografía.

original y la imagen eliminada de ruido. Mientras más bajo sea el valor del MSE, más alto es el valor PSNR y menor será el error.

Mientras tanto MSSIM se diferencia de estas técnicas porque es un modelo basado en la percepción visual que considera la degradación de la imagen como un cambio percibido en la información estructural, incluyendo el enmascaramiento de la luminancia y el enmascaramiento del contraste. Se utiliza el índice SSIM medio para evaluar la calidad general de la imagen.

Para obtener la medida PSNR y MSSIM se realizaron 5 experimentos por cada tipo de resolución (dando un total 10 experimentos) y en cada experimento elegimos de forma aleatoria como imágenes de prueba solo el 40 % de las imágenes de la base de datos, debido a que el 60 % restante es usado para la etapa de entrenamiento para el método basado en NMF. Tras haber contaminado las imágenes de prueba con los diferentes tipos y niveles de ruido por experimento, se procede a encontrar los valores PSNR y MSSIM cuando estas imágenes son y no son procesadas (imágenes ruidosas) por los diferentes métodos de eliminación de ruido. Posteriormente, se calcula el valor promedio de PSNR y MSSIM de todos los experimentos por familia tipográfica.

5.2. Mediciones basadas en MOS

Se realizó una prueba experimental con observadores humanos con el objetivo de obtener valoraciones subjetivas de la calidad de las tres tipografías tras haber pasado por los diferentes métodos de eliminación de ruido. Tras estas pruebas obtenemos una puntuación MOS que caracteriza dicha valoración subjetiva de la calidad de las imágenes. Todo esto basado en la recomendación de la ITU-T P.910 [ITU08].

Nuestro método de prueba está basado en el método de Índice por Categorías de Degradación (DCR, Degradation Category Rating) ya que el observador puede comparar directamente la secuencia procesada con la secuencia de referencia. Y fusionándolo con el modo de Presentación Simultánea (SP, Simultaneous Presentation) podemos mostrar las secuencias de prueba y de referencia en el mismo monitor. Por lo tanto, en las secuencias de nuestro método de prueba además de la muestra de referencia (imagen original) se presentará otras 3 muestras de prueba, siendo cada una procesada por diferentes métodos de eliminación de ruido o contaminadas con

distintos ruidos digitales. Aclarando que en cada secuencia, las muestras de prueba son contaminadas con el mismo tipo y nivel de ruido. Un ejemplo de secuencia empleada durante la evaluación se observa en la Figura 5.2.

Este modus operandi nos brinda una serie de ventajas como la reducción de tiempo en la prueba y una evaluación más sencilla para los observadores. Según las recomendaciones de [ITU08] hay que tener las siguientes consideraciones: las secuencias deben mostrarse en un fondo gris al 50 %, la distancia de observación debe ser de $1H$ a $8H$ (siendo H la altura de la imagen) y un monitor como mínimo de 14”.

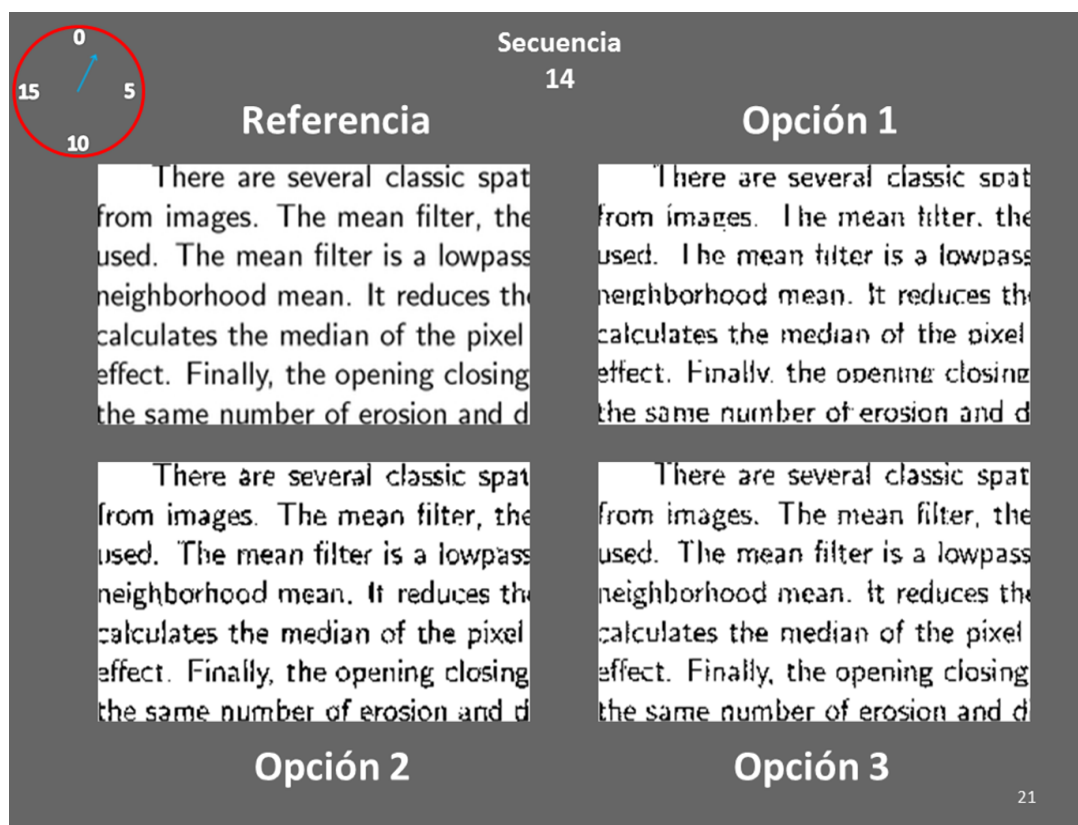


Figura 5.2: Ejemplo de secuencia, conformada por una muestra de referencia y tres de prueba, cada una procesada por diferente método de eliminación de ruido. Todas las opciones bajo las mismas condiciones.

Cada secuencia tendrá una duración 20 segundos, en donde los observadores durante ese lapso de tiempo tendrán que evaluar según su percepción la degradación de cada uno de las muestras de prueba en comparación a la muestra referente. En la esquina superior izquierda de cada secuencia (Figura 5.2) se encuentra un cronómetro para que los observadores tengan una noción del tiempo y puedan calificar antes de que la secuencia cambie. Previamente, a los sujetos de prueba se les explicó y entregó impreso (Tabla 5.2) la escala de medición de cinco niveles con la que trabajarán

durante la prueba. Dicha escala le permitirá al observador calificar que tanto afecta el proceso de eliminación de ruido o el ruido a la calidad de la imagen original en comparación a la imagen de prueba.

Calificación	Descripción
5	Imperceptible el cambio: No se percibe cambios entre la imagen original y la imagen de prueba, es decir, son iguales.
4	Perceptible pero no incómodo: Percibe algunos cambios entre la imagen original y la imagen de prueba pero no le afecta en la lectura.
3	Un poco incómodo: Los cambios hacen que la lectura sea un poco incómoda.
2	Incómodo: Los cambios dificultan aún más la lectura.
1	Muy incómodo: Los cambios son tan severos que imposibilita la lectura totalmente.

Cuadro 5.2: Escala de medición usada para la medición MOS.

5.2.1. Diseño experimental para la medida MOS

Según la recomendación de la ITU-T P.910, el número de observadores puede variar entre 15 a 40 (no se obtiene mayor ventaja si hay más de 40 observadores). Estas personas no deben intervenir directamente en la realización de las evaluaciones y no deben ser evaluadores experimentados.

Nuestro diseño experimental se dividió en 2 partes: una para las imágenes procesadas con los distintos métodos de eliminación de ruido y otro netamente para las imágenes ruidosas.

a) Imágenes procesadas

Primero seleccionamos 16 imágenes de referencia por tipo de tipografía (8 de resolución 280×530 y 8 de resolución 420×530), dando un total de 48 imágenes. Cada observador tendrá por prueba subjetiva 4 imágenes de referencia (que contienen 2 familias tipografías) de las 48 posibles. Es por ello que se tuvo que crear 12 pruebas subjetivas diferentes para que todas las imágenes sean evaluadas. Cada prueba subjetiva fue realizada en diferentes archivos creados usando Power Point.

Cada una de las 48 muestras de referencia son contaminadas por 3 tipos de ruido cuya potencia de ruido variará entre $0,01 \leq \sigma^2 \leq 0,1$ en pasos de 0.01 y serán procesadas por 3 métodos de eliminación de ruido. Es decir, si hacemos una combinación de estos tres elementos, de una imagen de referencia se generaran 90 muestras procesadas; y como son 48 imágenes de muestra en total tendremos 4320 muestras procesadas. Lo que equivale a 360 muestras procesadas por prueba subjetiva, que serán distribuidas

correctamente en 120 secuencias. En donde cada secuencia se conformará por una muestra de referencia y otras 3 muestras procesadas por los diferentes métodos de eliminación de ruido, que además se encuentran bajo el mismo tipo y nivel de ruido, como se muestra en la Figura 5.2.

b) Imágenes ruidosas

Al igual que el caso anterior también seleccionamos las mismas 16 imágenes de referencia por familia tipográfica, obteniendo un total de 48 imágenes. De una imagen de referencia se crearán 30 imágenes ruidosas, que se generaron a través de la combinación de los 3 tipos de ruido y su rango de varianzas de $0,01 \leq \sigma^2 \leq 0,1$ en pasos de 0.01. Por lo que en total tendremos 1440 imágenes ruidosas. Para que todas estas imágenes ruidosas sean calificadas, se creó 6 pruebas subjetivas que contienen 8 imágenes de referencia (con 2 familias tipográficas) cada una. En cada prueba subjetiva habrá 240 imágenes ruidosas que serán correctamente distribuidas en 80 secuencias. Un ejemplo de la estructura de cada secuencia la observamos en Figura 5.2.

Ya conociendo como se conformará cada prueba subjetiva en ambos casos (imágenes procesadas e imágenes ruidosas), hay que recordar que nuestro análisis va enfocado a las diferentes familias tipografías; y para poder acercarnos o superar el máximo número de observadores recomendado, en nuestra experimentación elegimos 48 personas. Por lo tanto, cada prueba subjetiva será utilizada por 4 observadores; lo que nos generará un total de 64 observadores por familia tipográfica.

5.2.2. Protocolo experimental para la obtención de la medida MOS

Las diversas pruebas subjetivas fueron creadas en diferentes archivos Power Point, donde se añadieron a las secuencias de prueba otras 2 secuencias de entrenamiento al comienzo, las cuales no se tuvieron en cuenta en el análisis estadístico de los resultados, y que nos servirán para que los observadores ajusten su opinión de acuerdo a la escala de medición (Tabla 5.2). Cada secuencia está programada para durar 20 segundos pero el observador tiene la potestad de pasar a la siguiente secuencia si es que terminó de evaluar la secuencia antes del tiempo mencionado. Por lo tanto, la evaluación de cada prueba subjetiva está diseñada para durar aproximadamente 25 minutos para las pruebas que contienen imágenes ruidosas y 40 minutos las pruebas que contienen imágenes procesadas.

En cuanto a la elección de los observadores, se pidió la ayuda de familiares y amigos ser partícipes del experimento para poder alcanzar el número propuesto de 48 observadores. Si alguno de ellos presentaba problemas con su vista se les pidió que utilicen sus lentes de medida. La edad de los observadores varía entre los 15 a 65 años. Para la votación, a los observadores se les entregó unas hojas de calificación (Figura 5.3), que junto a la escala de medición (Tabla 5.2) que se les entregó y explicó anteriormente, tendrán que calificar de acuerdo a su percepción y criterio cada muestra de prueba en comparación a la muestra referencial para cada secuencia.

Número de Secuencia	Opción 1	Opción 2	Opción 3
Secuencia 1			
Secuencia 2			
Secuencia 3			
Secuencia 4			
Secuencia 5			
Secuencia 6			
Secuencia 7			
Secuencia 8			
Secuencia 9			
Secuencia 10			

Figura 5.3: Una sección de las hojas de calificación.

5.2.3. Cálculo MOS

Tras haber obtenido los resultados de las hojas de calificación de las 48 personas, se procede a calcular los valores medios haciendo uso de la Puntuación Media de Opinión (MOS) que se obtiene de la siguiente manera

$$MOS = \frac{1}{N} \sum_{i=0}^N u_{ijkr} \quad (5.1)$$

donde u_{ijkr} es la votación de cada muestra procesada del observador i , para la familia tipográfica j , del tipo de ruido k y nivel de ruido r ; mientras que N es la cantidad de personas que votaron por familia tipográfica, es decir, 64 observadores. La medida MOS se calcula para cada método de eliminación de ruido en imágenes (basadas en wavelet, curvelet y NMF).

5.3. Resultados experimentales

En esta subsección se muestran los resultados experimentales llevados a cabo considerando tres diferentes tipos de tipografía (con serifa, sin serifa y máquina) usando dos medidas de evaluación objetiva (PSNR y MSSIM) y la medida subjetiva (MOS).

5.3.1. Tipografía con serifa

5.3.1.1. Resultados PSNR

En la Figura 5.4 se puede mostrar claramente que sin importar el tipo de ruido que contamine a las imágenes de documento que contienen tipografía con serifa, el método basado en curvelet presenta el mejor rendimiento. Solo en el caso de ruido impulsivo (Figura 5.4b) curvelet es similar con respecto a wavelet cuando la varianza de ruido es menor a 0.02. NMF por su parte es ligeramente superior a wavelet al momento de eliminar ruido gaussiano ($0,03 < \sigma^2 < 0,08$) y ruido multiplicativo ($0,03 < \sigma^2 < 0,09$) como se puede observar en la Figura 5.4a y Figura 5.4c, respectivamente. En cuanto a la eliminación de ruido impulsivo, wavelet obtiene casi los mismos valores que la línea base y NMF muestra ser superior a wavelet a partir de una varianza mayor a 0.02 (Figura 5.4b).

5.3.1.2. Resultados MSSIM

Cuando las imágenes son contaminadas con ruido gaussiano (Figura 5.5a) y ruido multiplicativo (Figura 5.5c), el método basado en curvelet obtiene los mejores resultados a lo largo de todo el rango de varianzas; mientras que NMF logra superar a wavelet a partir de una varianza mayor a $\sigma^2 > 0,03$. En la Figura 5.5b se puede observar que curvelet logra ser el método predominante a partir de una varianza mayor a 0.04 ya que entre un rango de $0,01 < \sigma^2 < 0,04$ el método basado en NMF es el que mejores resultados MSSIM presenta y wavelet es ligeramente superior cuando la varianza es muy pequeña $\sigma^2 = 0,01$.

5.3.1.3. Resultados MOS

En la Figura 5.6a, muestra que los métodos basados en curvelet y NMF no presenta grandes diferencias, ya que intercalan una superioridad entre ellos a lo largo del rango de varianzas de ruido. Sin embargo, ambos métodos son superiores al método basado en wavelet a partir de $\sigma^2 > 0,03$. En la Figura 5.6c, el método basado en NMF presenta mejores resultados que los otros dos métodos a partir de una varianza mayor a 0.02 y el método basado en curvelet es mejor a wavelet a partir de $\sigma^2 > 0,03$. Un resultado totalmente distinto vemos cuando las imágenes son contaminadas con ruido impulsivo (Figura 5.6b), en donde wavelet es el que mejor puntuación de opinión recibe; NMF logra superar al método basado en curvelet entre $0,02 < \sigma^2 < 0,08$.

Un breve resumen de los resultados de las imágenes con tipografía con serifa se puede observar en la Tabla 5.3.

	Gaussiano	Impulsivo	Multiplicativo
PSNR	-Curvelet mejor rendimiento. -NMF ligeramente superior a wavelet entre $0,03 < \sigma^2 < 0,08$.	-Curvelet mejor rendimiento $\sigma^2 > 0,04$. -NMF mayor a wavelet $\sigma^2 > 0,04$.	-Curvelet mejor rendimiento. -NMF ligeramente superior a wavelet entre $0,03 < \sigma^2 < 0,09$.
MSSIM	-Curvelet mejor rendimiento. -NMF superior a wavelet $\sigma^2 > 0,03$.	-Curvelet superior $\sigma^2 > 0,04$. -NMF superior a wavelet a partir de $\sigma^2 > 0,02$. -NMF superior a curvelet entre $0,02 < \sigma^2 < 0,04$.	-Curvelet mejor rendimiento. -NMF superior a wavelet $\sigma^2 > 0,04$.
MOS	-No hay diferencia significativa entre NMF y curvelet. -NMF y curvelet superior a wavelet $\sigma^2 > 0,03$.	-Wavelet mejor rendimiento. -NMF superior a curvelet entre $0,02 < \sigma^2 < 0,08$.	-NMF mejor rendimiento $\sigma^2 > 0,02$. -Curvelet superior a wavelet $\sigma^2 > 0,03$.

Cuadro 5.3: Tabla resumen de los resultados mostrados con la tipografía con serifa.

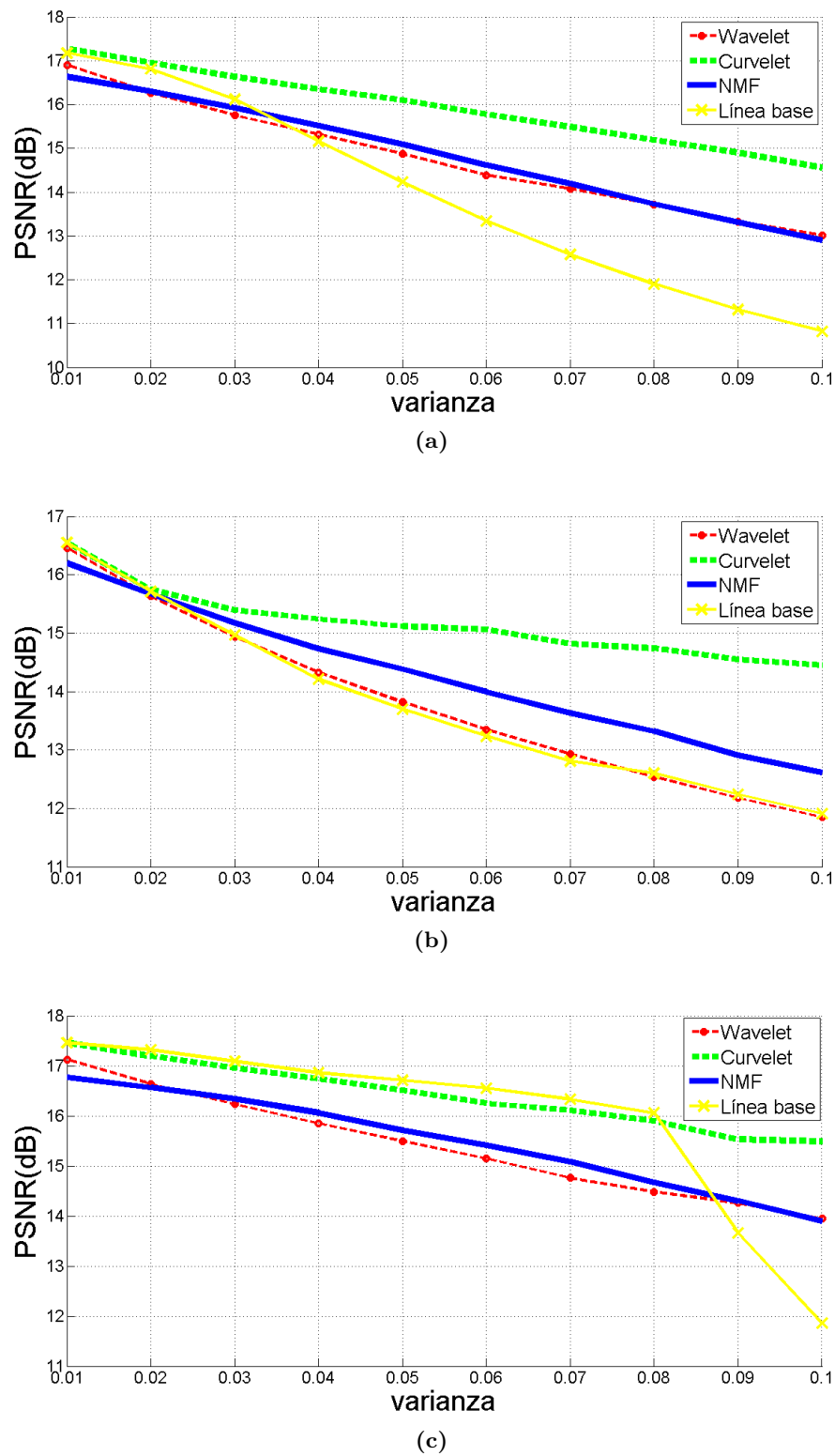
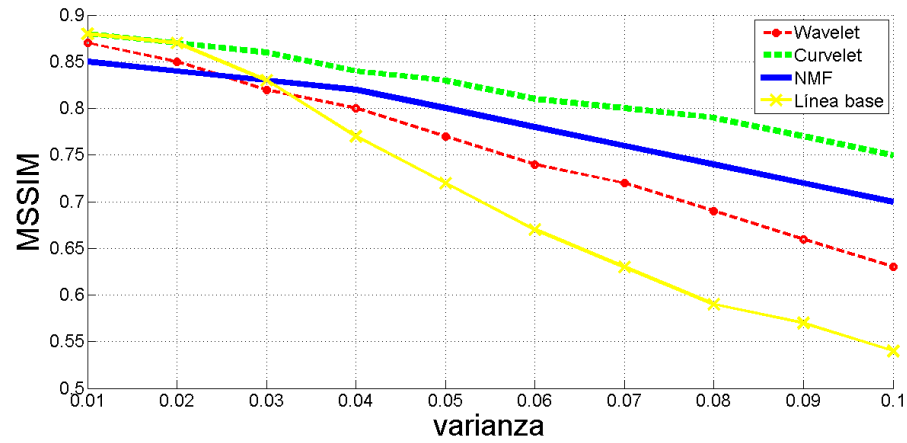
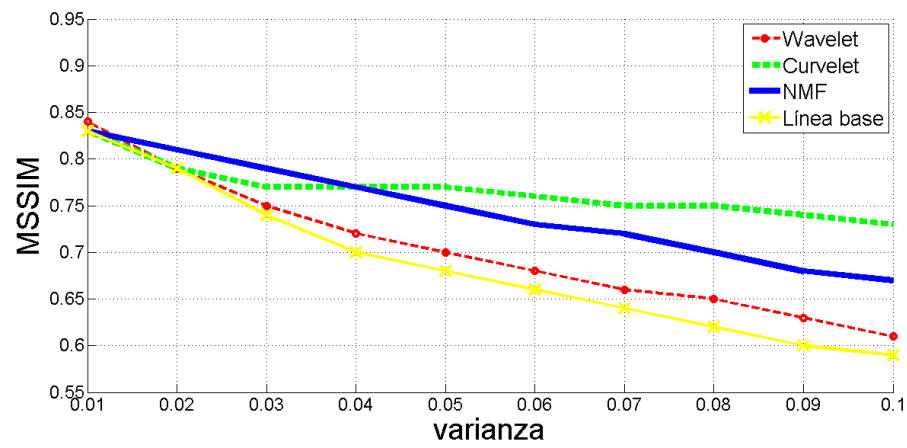


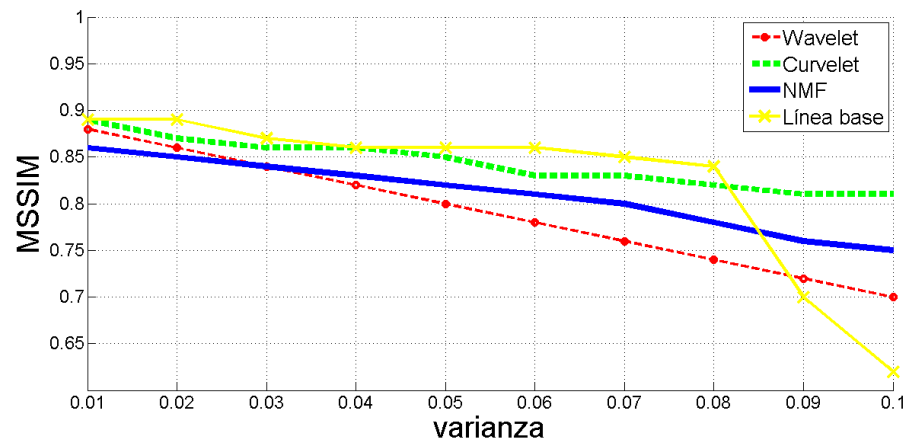
Figura 5.4: Resultados PSNR para tipografías con serifa contaminadas con ruido:
a) gaussiano, b) impulsivo y c) multiplicativo.



(a)



(b)



(c)

Figura 5.5: Resultados MSSIM para tipografías con serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.

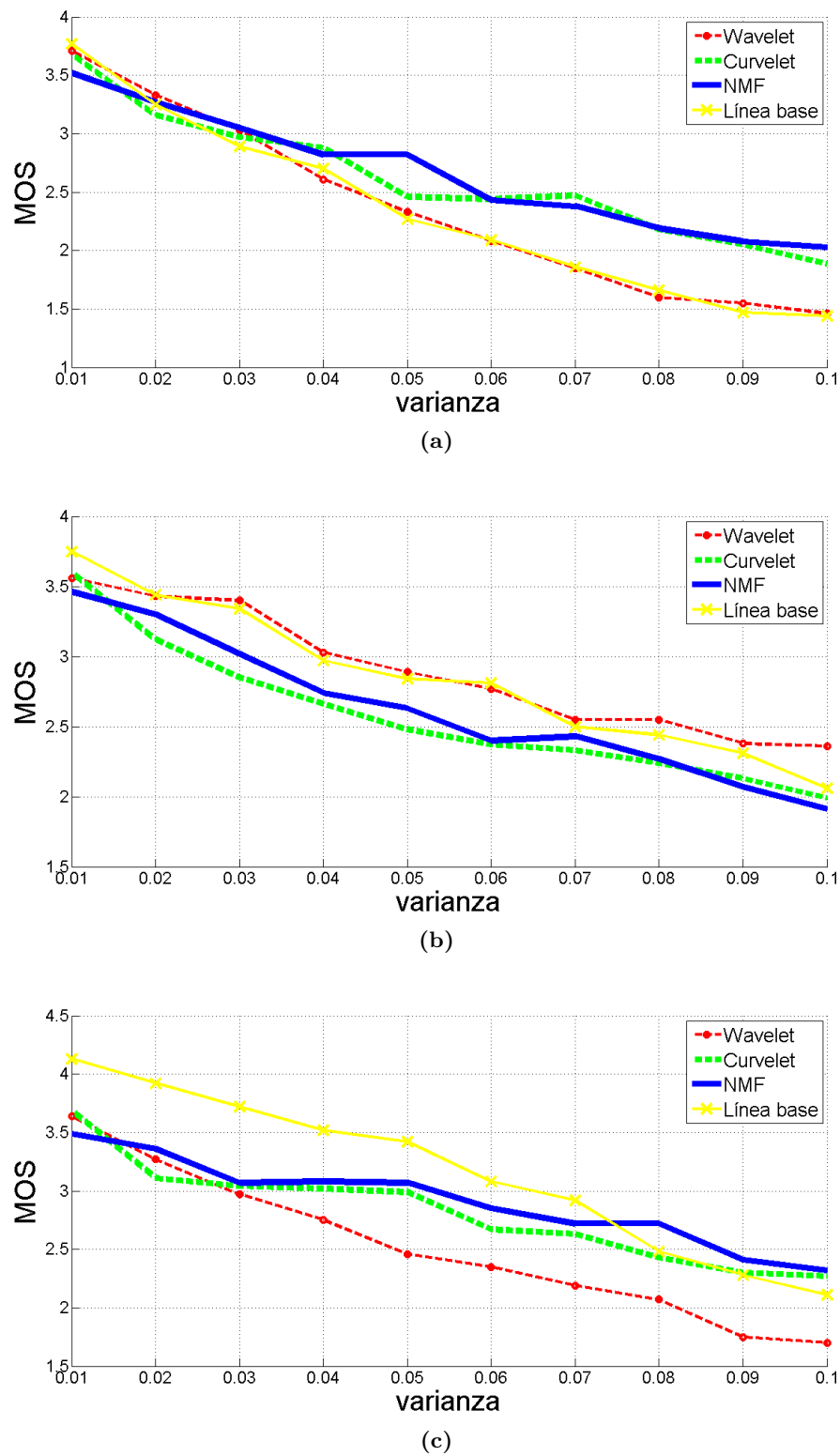


Figura 5.6: Resultados MOS para tipografías con serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.

5.3.2. Tipografía sin serifa

5.3.2.1. Resultados PSNR

En la Figura 5.7a y Figura 5.7c, se puede observar claramente que el método de eliminación de ruido basado en curvelet es que mejor desempeño presenta al momento de eliminar ruido gaussiano y multiplicativo, respectivamente. Mientras que para la eliminación de ruido impulsivo (Figura 5.7b) solo logra obtener mejores resultados de PSNR que los otros dos métodos con una varianza mayor a 0.05. Los resultados basados en NMF, este método logra superar ampliamente a wavelet con cualquier tipo de ruido excepto cuando son menores a $\sigma^2 > 0,03$, mientras que con ruido impulsivo (Figura 5.7b) NMF es ligeramente mejor que curvelet con una varianza de ruido entre $0,02 < \sigma^2 < 0,04$. Por otro lado, se puede observar que wavelet no logra eliminar el ruido impulsivo ya que obtiene casi los mismos valores de PSNR que los resultados de las imágenes contaminadas con este tipo de ruido (línea base).

5.3.2.2. Resultados MSSIM

Los resultados del método basado en NMF al momento de eliminar ruido gaussiano (Figura 5.7a) y multiplicativo (Figura 5.7c), muestran que este método logra ser superior a wavelet a partir de $\sigma^2 > 0,03$ y va ampliando su diferencia a medida que la varianza de ruido va aumentando; pero en ninguno de los dos casos logra superar al método basado en curvelet. En cuanto a la eliminación de ruido impulsivo (Figura 5.7b) los resultados MSSIM en el método basado en NMF son mayores a curvelet solo entre $0,01 < \sigma^2 < 0,05$ y es ampliamente superior al método basado en wavelet.

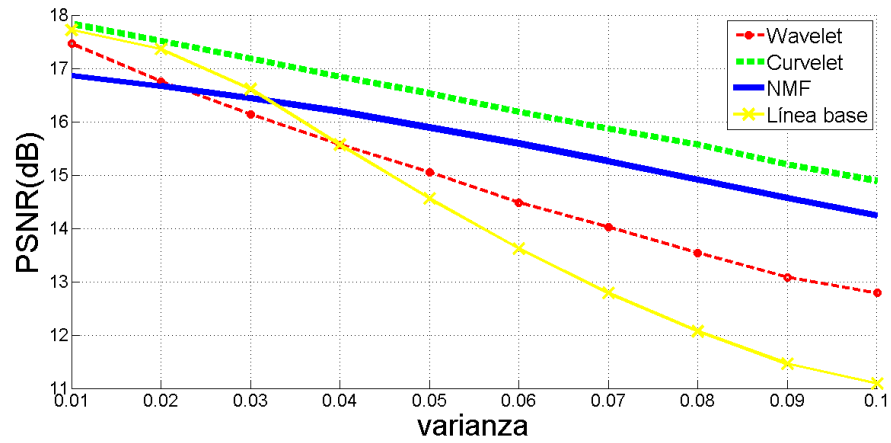
5.3.2.3. Resultados MOS

El método basado en NMF según los resultados MOS logra ser mejor que wavelet a partir de una varianza de ruido mayor a 0.03; sin embargo, el método basado en curvelet es el que mejores resultados MOS presenta a lo largo de todas las varianzas (Figura 5.9a). En cuanto a los resultados con ruido impulsivo (Figura 5.9b), el método de eliminación de ruido predominante es el basado en wavelet seguido por el método basado en curvelet; el método basado en NMF presenta resultados muy bajos. En la Figura 5.9c, curvelet es el método con mejores resultados a lo largo del rango de varianzas de ruido mientras que NMF logra superar a wavelet a partir de una varianza de ruido a partir de $\sigma^2 > 0,04$.

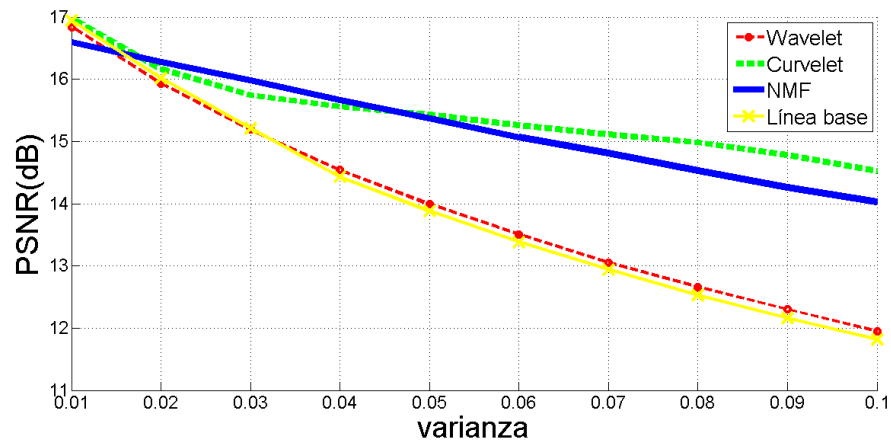
Un breve resumen de los resultados de las imágenes con tipografía sin serifa se puede observar en la Tabla 5.4.

	Gaussiano	Impulsivo	Multiplicativo
PSNR	-Curvelet mejor rendimiento. -NMF mejor a wavelet a partir de $\sigma^2 > 0,03$.	-Curvelet superior a partir de $\sigma^2 > 0,05$. -NMF mayor a wavelet $\sigma^2 \geq 0,02$.	-Curvelet mejor rendimiento. -NMF mejor a wavelet a partir de $\sigma^2 > 0,03$.
MSSIM	-Curvelet mejor rendimiento. -NMF mejor a wavelet a partir de $\sigma^2 > 0,03$.	-NMF superior a curvelet entre $0,01 < \sigma^2 < 0,05$ y superior a wavelet en todo el rango de varianzas.	-Curvelet mejor rendimiento. -NMF mejor a wavelet a partir de $\sigma^2 > 0,03$.
MOS	-Curvelet mejor rendimiento. -NMF superior a wavelet $\sigma^2 > 0,04$.	-Wavelet mejor rendimiento. -NMF muy bajo rendimiento.	-Curvelet mejor rendimiento. -NMF superior a wavelet $\sigma^2 > 0,04$.

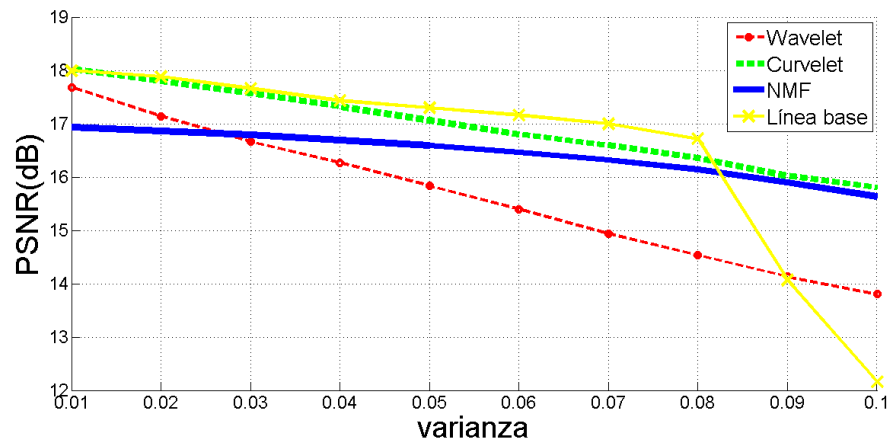
Cuadro 5.4: Tabla resumen de los resultados mostrados con la tipografía sin serifa.



(a)

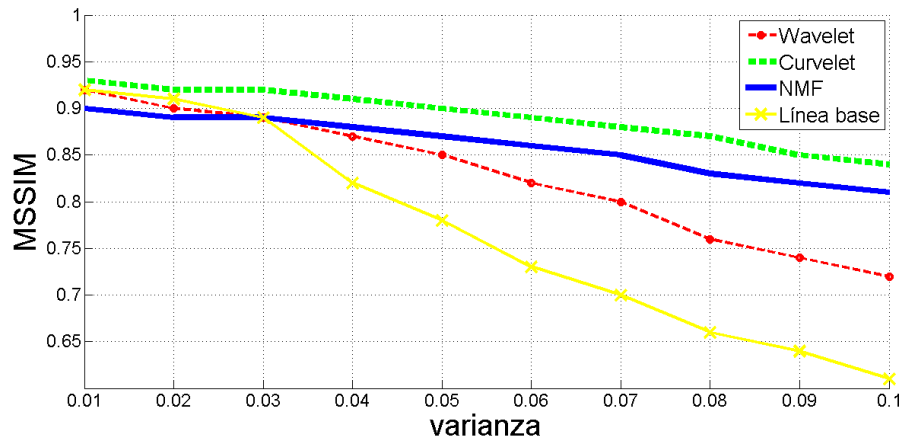


(b)

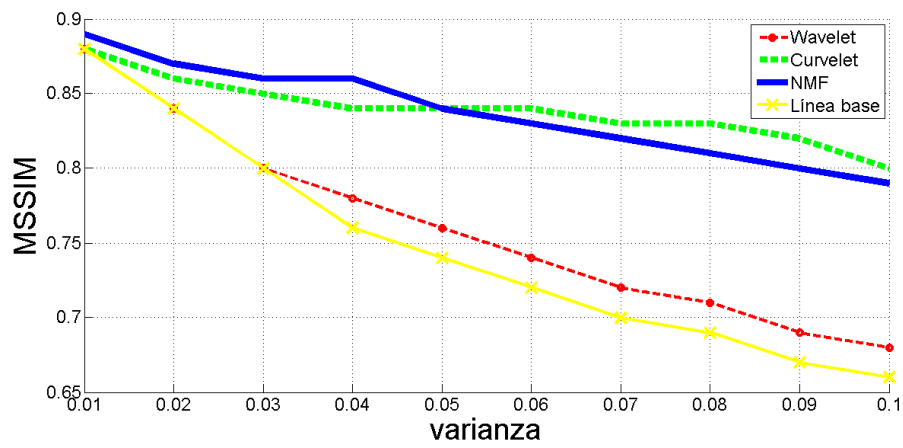


(c)

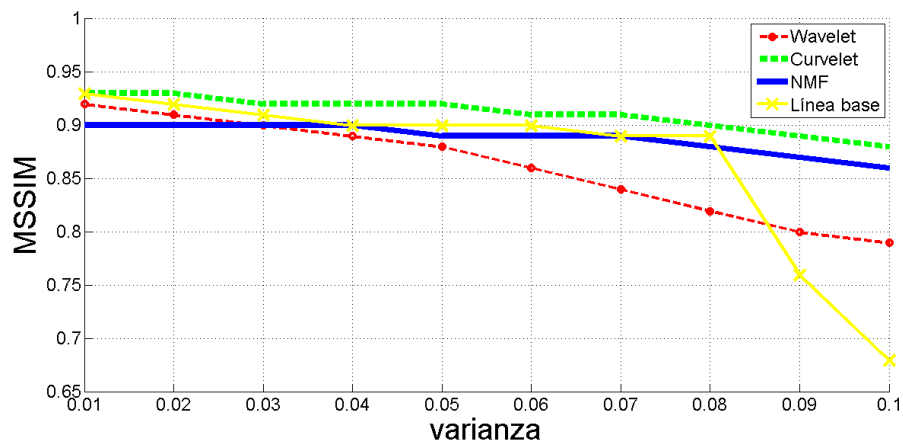
Figura 5.7: Resultados PSNR para tipografías sin serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.



(a)



(b)



(c)

Figura 5.8: Resultados MSSIM para tipografías sin serifa contaminadas con ruido:
a) gaussiano, b) impulsivo y c) multiplicativo.

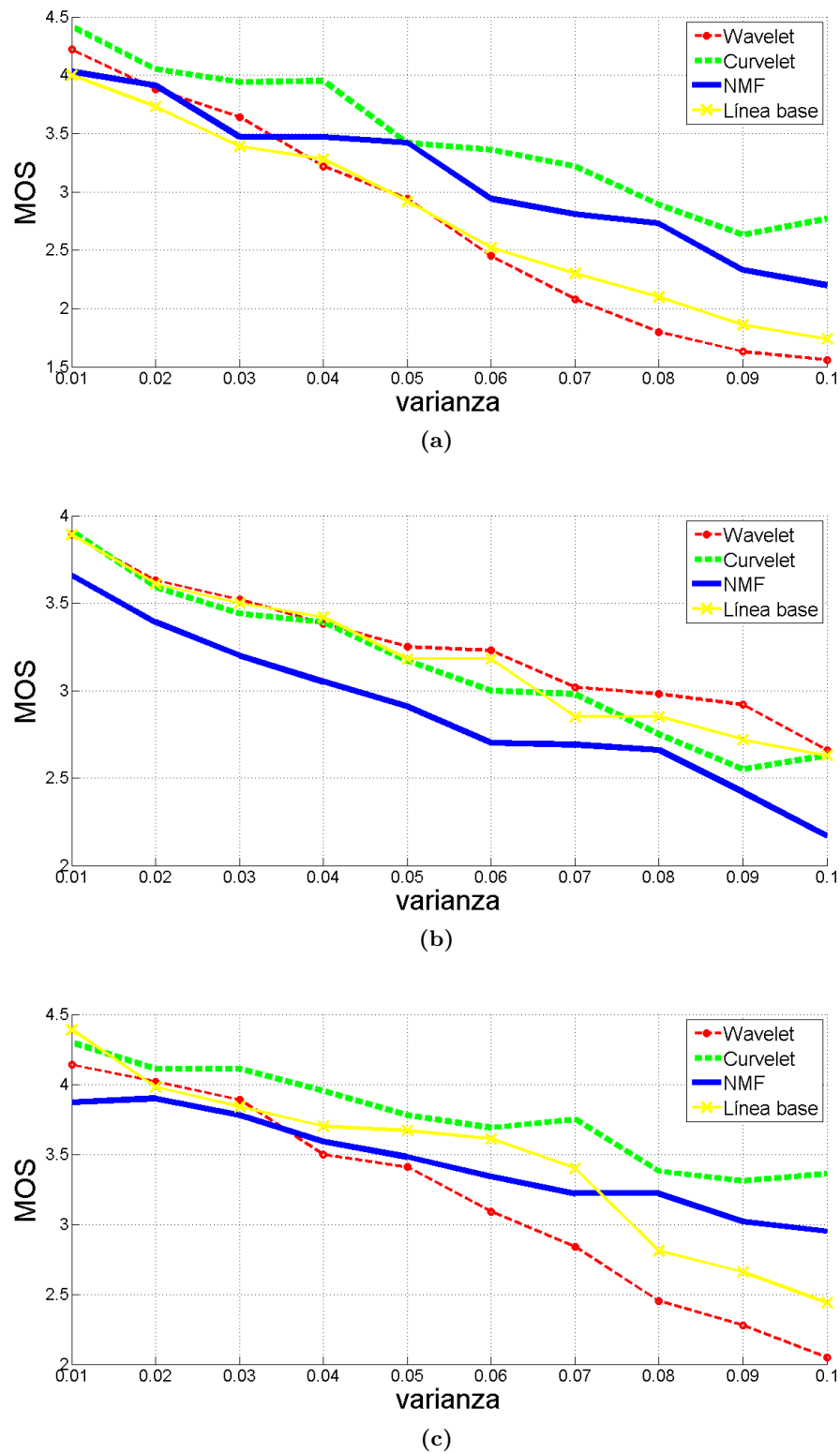


Figura 5.9: Resultados MOS para tipografías sin serifa contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.

5.3.3. Tipografía máquina de escribir

5.3.3.1. Resultados PSNR

En la Figura 5.10 es evidente que sin importar el tipo de ruido con que sea contaminada este tipo de tipografía los resultados PSNR del método basado en NMF son muy inferiores a wavelet y curvelet; además, curvelet muestra una gran superioridad en resultados. Solo en imágenes contaminadas con ruido impulsivo no existe una diferencia significativa entre curvelet y wavelet cuando la varianza de ruido es menor a 0.02, como se observa en la Figura 5.10b.

5.3.3.2. Resultados MSSIM

En el caso de los resultados de las mediciones MSSIM, el método basado en curvelet es superior a los otros dos métodos al momento de eliminar ruido gaussiano (Figura 5.10a) y ruido multiplicativo (Figura 5.10c), mientras que al eliminar ruido impulsivo (Figura 5.11) curvelet obtiene resultados casi iguales que wavelet entre $0,01 \leq \sigma^2 < 0,02$. Además, sigue existiendo la mala performance del método basado en NMF para eliminar cualquier tipo de ruido con la fuente del tipo máquina de escribir, como se observa en Figura 5.11.

5.3.3.3. Resultados MOS

En la Figura 5.12, la opinión de la gente nos confirma el mal rendimiento de NMF al eliminar el ruido en imágenes de documentos que presentan esta tipografía con cualquier tipo de ruido. Además se puede observar que el método basado en curvelet tiene una buena puntuación por parte de los observadores al momento de eliminar los ruidos gaussiano, impulsivo y multiplicativo. Solo con el ruido impulsivo, el método basado en wavelet logra acercarse al rendimiento del método basado en curvelet (Figura 5.12b).

Un breve resumen de los resultados de las imágenes con tipografía máquina de escribir se puede observar en la Tabla 5.5.

	Gaussiano	Impulsivo	Multiplicativo
PSNR	-Curvelet mejor rendimiento. -NMF muy bajo rendimiento.	-Curvelet superior a wavelet a partir de $\sigma^2 \geq 0,02$. -NMF muy bajo rendimiento.	-Curvelet mejor rendimiento. -NMF muy bajo rendimiento.
MSSIM	-Curvelet mejor rendimiento. -NMF muy bajo rendimiento.	-Curvelet superior a wavelet a partir de $\sigma^2 \geq 0,02$. -NMF muy bajo rendimiento.	-Curvelet mejor rendimiento. -NMF muy bajo rendimiento.
MOS	-Curvelet mejor rendimiento. -NMF muy bajo rendimiento.	-Curvelet mejor rendimiento. -NMF muy bajo rendimiento.	-Curvelet mejor rendimiento. -NMF muy bajo rendimiento.

Cuadro 5.5: Tabla resumen de los resultados mostrados con la tipografía máquina de escribir.

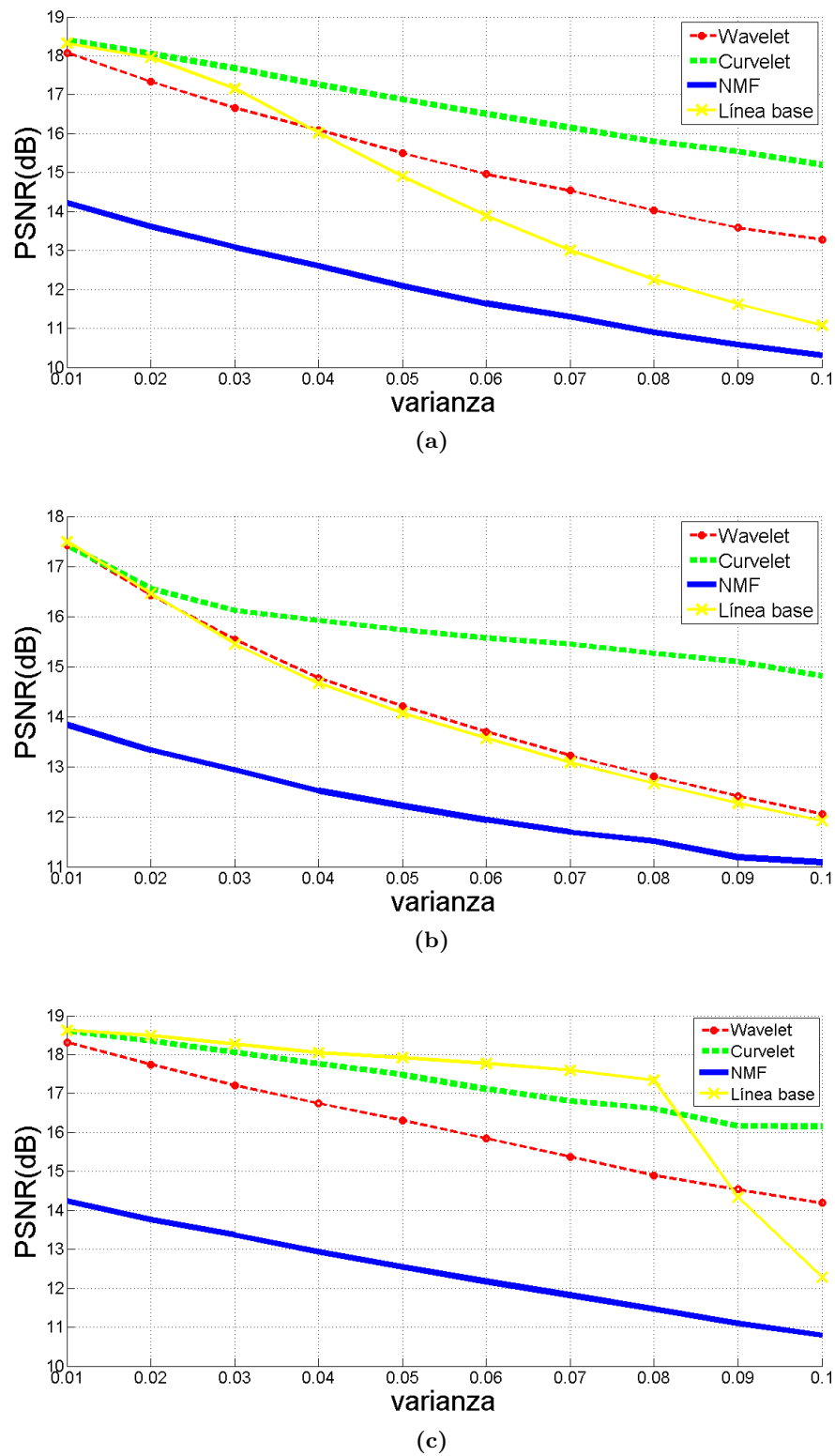
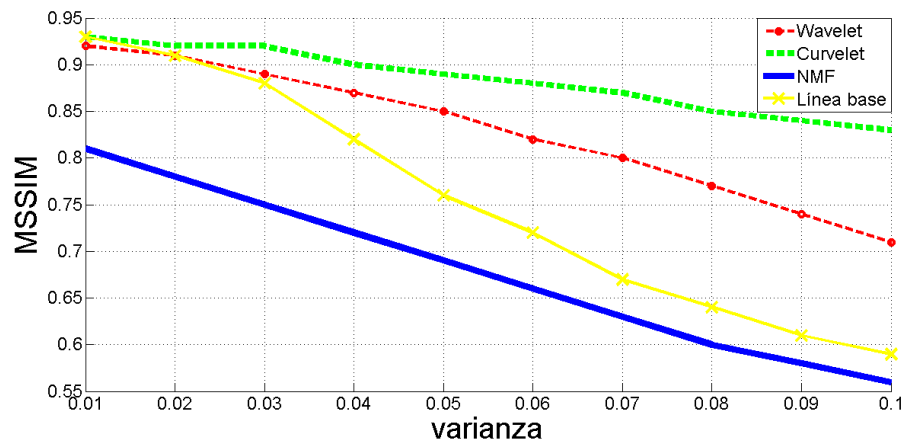
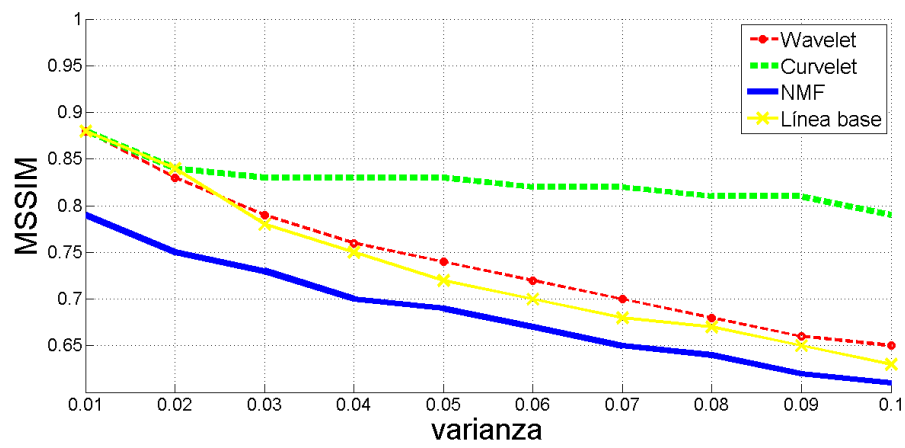


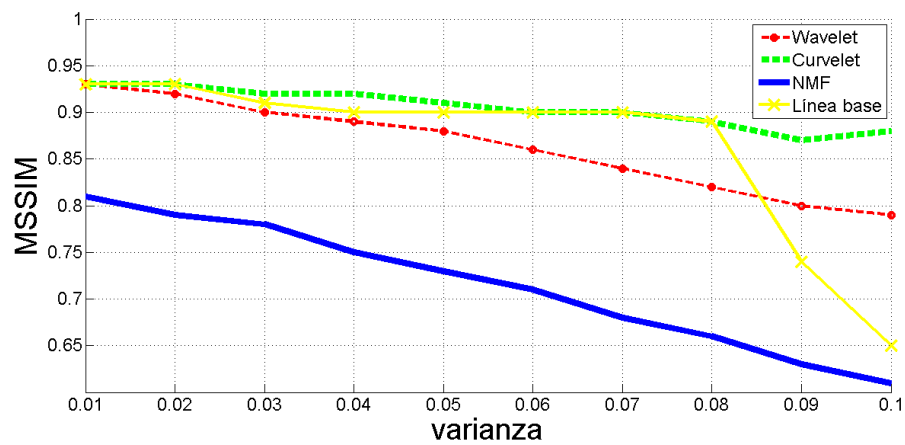
Figura 5.10: Resultados PSNR para tipografías maquina de escribir contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.



(a)



(b)



(c)

Figura 5.11: Resultados MSSIM para tipografías maquina de escribir contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.

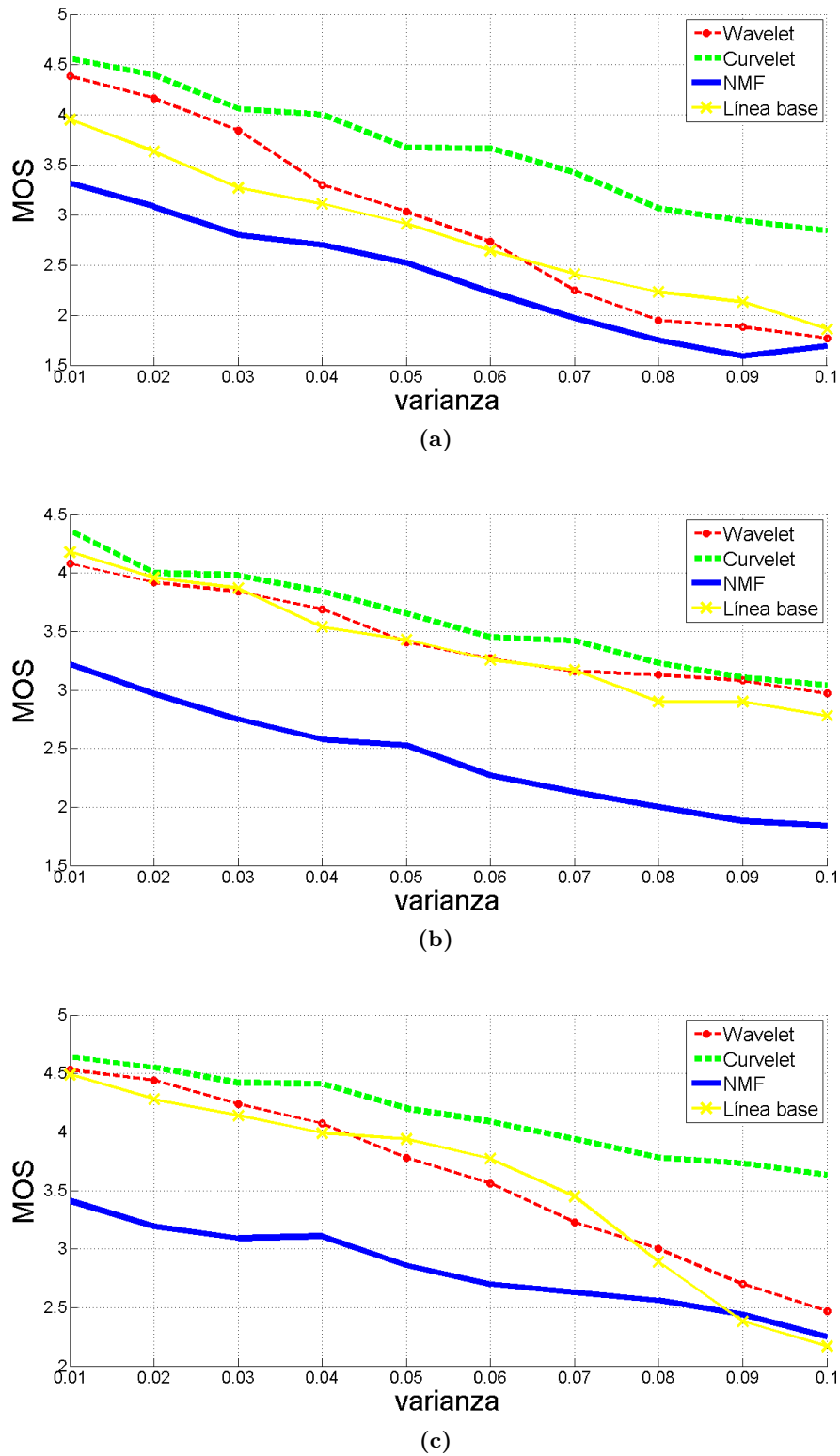


Figura 5.12: Resultados MOS para tipografías maquina de escribir contaminadas con ruido: a) gaussiano, b) impulsivo y c) multiplicativo.

5.3.4. Análisis comparativo

A continuación, se mostrará tablas que reflejan el promedio de los resultados PSNR, MSSIM y MOS de las distintas varianzas de ruido con que fueron contaminadas las imágenes de la base de datos. Esto nos ayudará a dar un análisis de la comparación de los resultados de las diferentes experimentaciones de forma más simple y concreta.

(a)

PSNR(dB)	Wavelet	Curvelet	NMF	Línea base
Gaussiano	14.76	15.92	14.82	13.70
Impulsivo	14.00	15.17	14.26	13.54
Multiplicativo	15.39	16.41	15.48	15.74

(b)

MSSIM	Wavelet	Curvelet	NMF	Línea base
Gaussiano	0.76	0.82	0.78	0.69
Impulsivo	0.70	0.77	0.75	0.68
Multiplicativo	0.79	0.84	0.81	0.80

(c)

MOS	Wavelet	Curvelet	NMF	Línea base
Gaussiano	2.35	2.62	2.66	2.34
Impulsivo	2.89	2.58	2.62	2.85
Multiplicativo	2.52	2.81	2.91	3.16

Cuadro 5.6: Tablas de promedio de las medidas de evaluación: a) PSNR, b) MSSIM y c) MOS, considerando las distintas varianzas de ruido que contaminan las imágenes con tipografía con serifa.

NMF demuestra que para las familias tipográficas con y sin serifa contaminadas con los ruidos puestos a prueba, sus resultados PSNR y MSSIM logran superar a wavelet; mientras que sus resultados MOS cuando son contaminadas con ruido gaussiano y multiplicativo supera al método basado en wavelet y solo es superior al método basado en curvelet con tipografía con serifa, según los resultados de las Tabla 5.6 y Tabla 5.7. Sin embargo, NMF no logra obtener un buen desempeño al eliminar el ruido en imágenes de documentos con tipografía máquina de escribir sin importar el tipo de ruido tanto para medidas objetivas como subjetiva, incluso siendo inferior a la línea base (Tabla 5.8).

(a)

PSNR(dB)	Wavelet	Curvelet	NMF	Línea base
Gaussiano	14.89	16.36	15.66	14.18
Impulsivo	13.99	15.45	15.26	13.95
Multiplicativo	15.64	16.93	16.43	16.49

(b)

MSSIM	Wavelet	Curvelet	NMF	Línea base
Gaussiano	0.83	0.89	0.86	0.75
Impulsivo	0.76	0.84	0.84	0.74
Multiplicativo	0.86	0.91	0.89	0.85

(c)

MOS	Wavelet	Curvelet	NMF	Línea base
Gaussiano	2.74	3.46	3.13	2.78
Impulsivo	3.25	3.14	2.88	3.18
Multiplicativo	3.17	3.77	3.44	3.29

Cuadro 5.7: Tablas de promedio de las medidas de evaluación: a) PSNR, b) MSSIM y c) MOS, considerando las distintas varianzas de ruido que contaminan las imágenes con tipografía sin serifa.

(a)

PSNR(dB)	Wavelet	Curvelet	NMF	Línea base
Gaussiano	15.41	16.75	12.03	14.57
Impulsivo	14.25	15.79	12.23	14.14
Multiplicativo	16.12	17.31	12.42	17.02

(b)

MSSIM	Wavelet	Curvelet	NMF	Línea base
Gaussiano	0.83	0.88	0.68	0.73
Impulsivo	0.74	0.83	0.68	0.72
Multiplicativo	0.86	0.91	0.72	0.84

(c)

MOS	Wavelet	Curvelet	NMF	Línea base
Gaussiano	2.93	3.66	2.36	2.81
Impulsivo	3.46	3.61	2.42	3.33
Multiplicativo	3.60	4.14	2.82	3.47

Cuadro 5.8: Tablas de promedio de las medidas de evaluación: a) PSNR, b) MSSIM y c) MOS, considerando las distintas varianzas de ruido que contaminan las imágenes con tipografía máquina de escribir.

Las comparaciones también nos manifiestan como NMF logra obtener su mejor desempeño en tipografías sin serifa (Tabla 5.7) debido a que no contiene los remates ubicados en los extremos de los caracteres como si lo tienen las tipografías con serifa y máquina de escribir. Estos remates se pixelean en imágenes digitales, en especial, si son de baja resolución provocando que sean confundidos como ruido, lo que genera que NMF al extraer las características significativas de la imagen (ruido) los remates también sean removidos deformando los caracteres. Como se puede observar en la Figura 5.13, a pesar de que NMF produce un ligero deterioro en los extremos de los caracteres de las letras sin serifa, no se compara al deterioro producido en las imágenes de las letras con serifa. Este inconveniente también lo sufren los otros dos métodos de eliminación de ruido en menor o mayor medida según el tipo e intensidad del ruido por lo que también obtienen sus mejores resultados con esta tipografía.

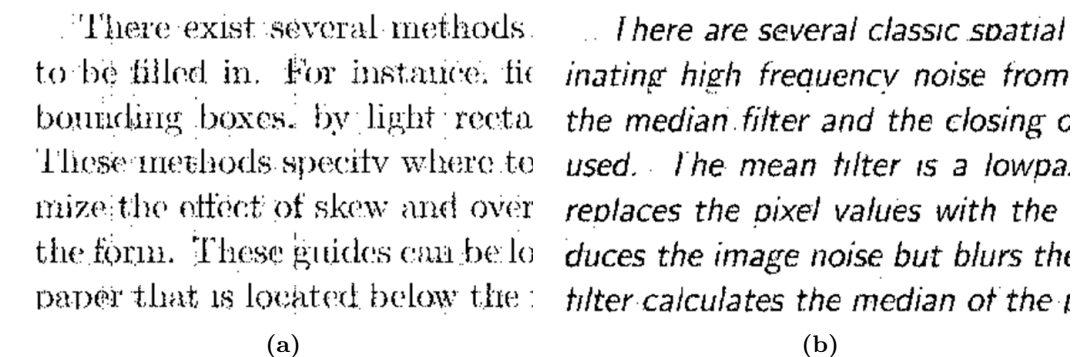


Figura 5.13: Ejemplos de imágenes de documentos tras el proceso de eliminación de ruido basado en NMF con tipografía: a) con serifa y b) sin serifa. Las imágenes fueron contaminadas con ruido gaussiano de media cero y varianza 0.05.

Según los resultados de las Tabla 5.6, Tabla 5.7 y Tabla 5.8, el método basado en curvelet nos muestra su superioridad basándonos en sus resultados PSNR y MSSIM para las tres familias tipográficas cuando han sido contaminadas con los distintos ruidos digitales. Esto se debe a que una evaluación PSNR es un análisis cuantitativo de que tan fidedigna es la imagen procesada con la original (bit por bit) y la evaluación MSSIM está basado en la percepción humana. Al fin y al cabo, ambas son medidas objetivas ya que están basadas en formulaciones matemáticas y solo se diferencian en el grado de correlación que presentan con respecto a la medida subjetiva. Sin embargo, con MOS se observa que hay un intercalamiento entre que método tiene el mejor desempeño en las distintas evaluaciones (excepto con tipografía máquina de escribir donde predomina curvelet). Esto ocurre ya que el criterio de valoración de calidad de los distintos observadores es subjetivo y no siempre coincidirá con los resultados de las mediciones objetivas ya que ellos basan su votación más que todo si es que se puede leer o no el texto y no en la igualdad exacta entre las imágenes

de referencia y de prueba.

Al momento de extraer los resultados provenientes de las mediciones MOS al eliminar ruido impulsivo, nos sorprendimos al ver las buenas calificaciones del método basado en wavelet a pesar de que sus mediciones objetivas digan lo contrario. Si bien los métodos basados en curvelet y NMF logran eliminar el ruido de fondo (los puntos negros), lamentablemente rompen los bordes de las letras, en especial NMF. En cambio, wavelet aunque no logra eliminar bien el ruido de fondo tampoco daña mucho las formas de los caracteres. Es por eso que recibe una mayor puntuación ya que se prioriza la legibilidad del texto. Un ejemplo de este caso lo podemos observar en la Figura 5.14.

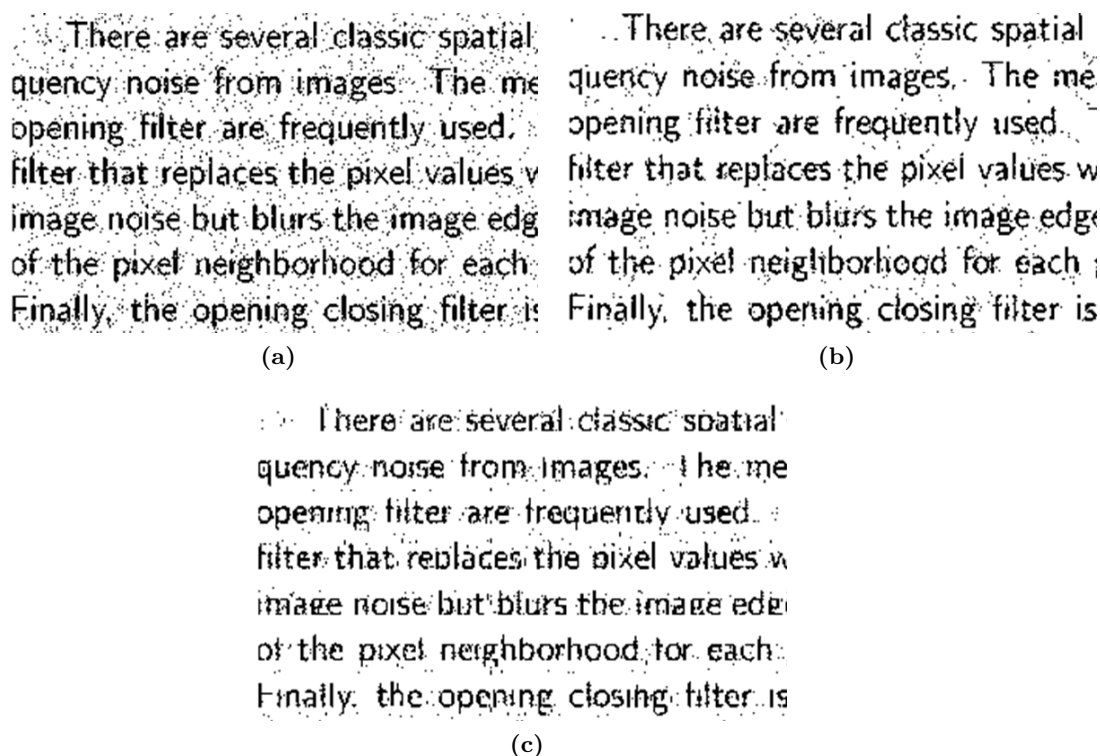


Figura 5.14: Ejemplos de imágenes de documentos de tipografía con serifa tras pasar por los métodos de eliminación de ruido basados en : a) wavelet, b) curvelet, c) NMF. Las imágenes fueron contaminadas con ruido impulsivo de media cero y varianza 0.06.

Otro caso curioso del método basado en wavelet sucede con los resultados MOS al momento de eliminar ruido gaussiano y multiplicativo. A pesar de que con las medidas PSNR y MSSIM los resultados de este método son siempre superiores a la línea base sin importar la tipografía, pero no sucede lo mismo con la medida subjetiva MOS. Porque la umbralización dura presenta desventaja en niveles altos

6 Conclusiones y trabajos futuros

6.1. Conclusiones

En este trabajo de tesis se ha presentado un estudio comparativo entre dos métodos convencionales de eliminación de ruido (basado en Wavelet y Curvelet) y el método basado en NMF aplicado a imágenes de documentos usando tres diferentes tipos de tipografías. En el caso del método basado en NMF se realizó control de la restricción de dispersión, usando la función de coste Kullback-Leibler. NMF es capaz de extraer características de la imagen que posteriormente nos servirán para eliminar el ruido en la imagen.

En la implementación del método de eliminación de ruido basado en la Transformada Wavelet, este falla al eliminar ruido gaussiano, impulsivo y multiplicativo cuando su varianza es alta debido a que las wavelets no son buenas en obtener representaciones direccionales de los bordes y por las desventajas de la umbralización dura al eliminar niveles altos de ruido. Por lo que este método necesita de otros post-procesamientos para mejorar su rendimiento a la hora de eliminar mejor ruido en imágenes que contengan texto.

Mientras que en la implementación del método basado en la Transformada Curvelet, obtiene un mejor rendimiento en las mediciones PSNR, MSSIM y MOS en la mayoría de las experimentaciones debido a que las curvelets presentan una mayor eficacia al momento de representar objetos con discontinuidades a lo largo de sus curvas, dándole la ventaja de preservar mejor las características de los bordes de los caracteres en las imágenes de documentos que los otros dos métodos puestos a prueba.

Por otra lado, nuestra implementación del método basado en NMF demuestra que al eliminar los ruidos puestos a prueba en imágenes de documentos con tipografía con serifa sus resultados PSNR y MSSIM son superiores al método basado en wavelet mas no al método basado en curvelet. Mientras que con los resultados de calidad MOS el método basado en NMF si logra superar los resultados del método basado en curvelet.

En lo que respecta a la eliminación de ruidos puestos a prueba en imágenes de documentos con tipografía sin serifa el método basado en NMF logra buenos resultados de calidad basados en PSNR y MSSIM ya que logra superar ampliamente al método basado en wavelet y presenta resultados cercanos al método basado en curvelet a

medida que el nivel de ruido aumenta. En cuanto a la puntuación dada por las personas, se confirma la superioridad de NMF con respecto a wavelet al eliminar ruido gaussiano y multiplicativo mas no con el ruido impulsivo. Cabe resaltar que NMF presenta su mejor rendimiento con esta tipografía.

Sin importar el tipo de ruido que se tenga que eliminar en imágenes de documentos con tipografía máquina de escribir, al implementar el método basado en NMF presenta un mal rendimiento usando las medidas PSNR, MSSIM y MOS ya que su rendimiento se encuentra muy por debajo de lo esperado. Esto se debe que tras el proceso de eliminación de ruido aparecen unas líneas verticales negras que van aumentando de número y densidad a medida que va creciendo la varianza de ruido. Este fenómeno solo sucede con las imágenes de documentos con tipografía máquina de escribir.

Finalmente, se puede concluir que a pesar que el método NMF no es exclusivo para la eliminación de ruido en imágenes y mucho menos en aquellas que contienen texto, se demostró que tiene potencial para tener ese uso, ya que este método supera en la mayoría de casos de mediciones objetivas al método basado en wavelet con tipografías con y sin serifa; y cuando la varianza va aumentado se va acercando al método basado en curvelet con tipografías sin serifa. Además, con la experimentación subjetiva se percibió como es que a las personas se les hace difícil encontrar diferencias considerables entre las imágenes procesadas por los diferentes métodos, provocando que se reduzca las diferencias estadísticas entre los diversos métodos al mostrar los resultados MOS en comparación con los resultados basados en PSNR y MSSIM.

6.2. Trabajos futuros

Como trabajo futuro se plantea mejorar el método de eliminación de ruido basado usando NMF para preservar mejor los bordes de los textos en imágenes de documentos, hasta llegar a poder eliminar ruido en imágenes de documentos con otras familias tipografías y caligrafías tanto en documentos actuales como históricos

Agradecimientos

Agradecer profundamente al Dr. Jimmy Diestín Ludeña Choez por su apoyo, paciencia, y guiarme de la mejor manera para alcanzar mis objetivos.

A mis padres, Antonio y Judith, por siempre apoyarme durante toda mi vida, sin su aliento y esfuerzo no estaría en estas etapas finales de mi carrera universitaria.

A todos mis amigos y compañeros, por sacarme sonrisas en tiempos difíciles y hacer que mi vida sea divertida con su compañía.

Gracias.

Bibliografía

- [AM14] R. Ansari and B. Mohan. Noise filtering of remotely sensed images using iterative thresholding of wavelet and curvelet transforms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Noviembre 2014.
- [Bal07] C. Balbuena. Metodo de proteccion con marca de agua para imagenes digitales utilizando una tecnica de representacion orientada geometricamente contourlet. *Universidad Nacional Autonoma de Mexico*, 2007.
- [BB05] M. Berry and M. Browne. Email surveillance using non negative matrix factorization. pages 249–264, 2005.
- [BDSH] A. Bertrand, K. Demuynck, V. Stouten, and H. Van hamme. Unsupervised learning of auditory filter banks using non-negative matrix factorization. *Proc. of the Acoustics, Speech and Signal Processing IEEE International Conference, ICASSP*, pages 4713–4716.
- [BF10] J. Bioucas and M. Figueiredo. Multiplicative noise removal using variable splitting and constrained optimization. *IEEE Transaction on Image Processing*, 2010.
- [BJ15] A. Boyat and B. Joshi. A review paper: Noise models in digital image processing. *Signal & Image Processing : An International Journal (SIPIJ)*, 6(2), Abril 2015.
- [BLM01] M. Bernard, C. Hui Liao, and M. Mills. The effects of font type and size on the legibility and reading time of online text by older adults. *Department of Psychology Wichita State University*, 2001.
- [BP04] I. Buciu and I. Pitas. Application of non-negative and local non negative matrix factorization to facial expression recognition. *ICPR*, pages 288–291, 2004.
- [Can10] J. Canonico. Compresion de imagenes medicas basada en transformada ondícula (wavelet) para aplicaciones en telemedicina. Master’s thesis, Escuela de Ingenieria de Telecomunicaciones , Universidad Catolica Andres Bello, Febrero 2010.
- [Cat91] E. Catich. *The Origin of the Serif: Brush Writings and Roman Letters*. Saint Ambrose University Catich Gallery, 1991.
- [Cat16] P. Cattin. Image restoration: Introduction to signal and image processing. *MIAC, Universidad de Basel*, Abril 2016.

- [CCR06] Z. Chen, A. Cichocki, and T. M. Rutkowski. Constrained non negative matrix factorization method for eeg analysis in early detection of alzheimers disease. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 893–896, 2006.
- [CD99] E. Candes and D. Donoho. Ridgelets: the key to higher-dimensional intermittency? 1999.
- [CD00] E. Candes and D. Donoho. Curvelets. a surprisingly effective nonadaptive representation for objects with edges. *Departamento de Estadísticas, Universidad de Stanford*, 2000.
- [CD02] E. Candes and L. Demanet. Curvelets and fourier integral operators. Noviembre 2002.
- [CDDY05] E. Candes, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *Applied and Computational Mathematics, Caltech, Pasadena, CA*, Julio 2005.
- [CDDY15] E. Candes, L. Demanet, D. Donoho, and L. Ying. Curvelab.org, Agosto 2015.
- [Che06] K. Cheng. *Disenar Tipografia*. 2006.
- [CLKC08] A. Cichocki, H.Y. Lee, Y.D. Kim, and S. Choi. Non-negative matrix factorization with alpha-divergence. *PRL*, (1433-1440), Julio 2008.
- [CZ06] A. Cichocki and R. Zdunek. Multilayer nonnegative matrix factorization. *Electronics Letters*, 42(16): 947–948., 2006.
- [CZA06a] A. Cichocki, R. Zdunek, and S. Amari. Csiszar’s divergences for non-negative matrix factorization: Family of new algorithms. *LNCS*, 3889: 32–39, 2006.
- [CZA06b] A. Cichocki, R. Zdunek, and S. Amari. New algorithms for non-negative matrix factorization in applications to blind source separation. *Proc. of the Acoustics, Speech, and Signal Processing*, 2006.
- [CZPA09] A. Cichocki, R. Zdunek, A. Huy Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations*. Wiley, 2009.
- [Dev05] K. Devarajan. Molecular pattern discovery using non-negative matrix factorization based on renyi’s information measure. In *Twelfth Annual International Conference on Statistics, Combinatorics, Mathematics and Applications*, 2005.
- [DJ94] D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 1994.
- [DMSS07] S. Du, X. Mao, P. Sajda, and D. C. Shungu. Automated tissue segmentation and blind recovery of 1h mrs imaging spectral patterns of normal and diseased human brain. *NMR in Biomedicine*, 2007.

- [Don95] D. Donoho. Denoising by soft thresholding. *IEEE Transaction on Information Theory*, 41, 1995.
- [DRdFC07] K. Drakakis, S. Rickard, R. de Frein, and A. Cichocki. Analysis of financial data using non-negative matrix factorization. *International Journal of Mathematical Sciences*, 2007.
- [DS05] I. S. Dhillon and S. Sra. Generalized nonnegative matrix approximations with bregman divergences. *Neural Information Processing Systems*, pages 283–290, 2005.
- [dSMS12] R. da Silva, R. Minetto, and W. R. Schwartz. *Adaptive edge preserving image denoising using wavelet Transforms*. Springer, 2012.
- [FBD09] C. Fevotte, N. Bertin, and J. Durrieu. Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis. *Neural Comput*, pages 793–830, 2009.
- [GC05] Y. Gao and G. Church. Improving molecular cancer class discovery through sparse non-negative matrix factorization. *Bioinformatics*, pages 3970–3975, 2005.
- [GE11] E. Grais and H. Erdogan. Single channel speech music separation using nonnegative matrix factorization and spectral masks. *Facultad de Ingeniería y Ciencias Naturales*, 2011.
- [Gir96] B. Girod. Whats wrong with mean-squared error? *Digital images and human vision (MIT Press)*, 1996.
- [GMC10] B. Gangamma, S. Murthy, and P. Chandra. A combined approach for degraded historical documents denoising using curvelet and mathematical morphology. *Department of ISE, PES Institute of Technology, Karnataka*, 2010.
- [GPH04] C. Gobinet, E. Perrin, and R. Huez. Application of nonnegative matrix factorization to fluorescence spectroscopy. In *Proc. EUSIPCO 2004*, Setiembre 2004.
- [GS13] B. Gupta and S. Singh. Image denoising with linear and non linear filters: A review. *IJCSI International Journal of Computer Science Issues*, 10(2), Noviembre 2013.
- [GW02] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, second edition edition, 2002.
- [Hoy04] P. Hoyer. Non negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, Noviembre 2004.
- [Hyv97] A. Hyvarinen. Sparse coding shrinkage: denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 1997.
- [Inc17] Kaggle Inc. The home of data science & machine learning. 2017, <https://www.kaggle.com/c/denoising-dirty-documents>.

-
- [ITU96] Methods for subjective determination of transmission quality. *ITU-T P.800*, 1996.
- [ITU06] Vocabulary for performance and quality of service. Technical report, ITU-T Rec. P.10, 2006.
- [ITU08] Subjective video quality assessment methods for multimedia applications. *ITU-T P.910*, 2008.
- [Jua] D. Alvarez Juarez. Introduccion a la tipografia. *Universidad de Londres*.
- [KM02] S. Kouro and R. Musalem. Tutorial introductorio a la teoria wavelet. 2002.
- [Kom07] R. Kompass. A generalized divergence measure for nonnegative matrix factorization. *Neural Comput*, pages 780–791, 2007.
- [LG12] J. Ludena and A. Gallardo. Speech denoising using non - negative matrix factorization with kullback - leibler divergence and sparseness constraints. *Advances in Speech and Language Technologies for Iberian Languages (IberSpeech2012)*, 2012.
- [LG15] J. Ludena and A. Gallardo. Contribuciones a la aplicacion de la factorizacion de matrices no negativas a las tecnologias del habla. Master’s thesis, Departamento de teoria de la senal y comunicaciones, Abril 2015.
- [LHZC03] S. Li, X. Hou, H. Zhang, and Q. Cheng. Learning spatially localized, partsbased representation. *Proceedings of the 2001 IEEE Computer Society Conference*, Abril 2003.
- [LLCL01] J. S. Lee, D. D. Lee, S. Choi, and D. S. Lee. Application of non-negative matrix factorization to dynamic positron emission tomography. *Proceedings of the ICA*, pages 629–632, 2001.
- [LS99] D. Lee and S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999.
- [LS01] D. Lee and H. Seung. Algorithms for non negative matrix factorization. *MIT Press*, 2001.
- [LZL04] W. Liu, N. Zheng, and X. Li. Nonnegative matrix factorization for eeg signal classification. *ISNN*, pages 470–475, 2004.
- [Mar02] J. Martinez. *Analisis de la teoria de ondas orientada a las aplicaciones en ingenieria electrica: Fundamentos*. 2002.
- [Mar13] M. Martin. *Procesamiento Digital de Imagenes*. Universidad Autonoma de Puebla, Enero 2013.
- [MP10] J. Ma and G. Plonka. The curvelet transform: A review of recent applications. *IEEE SIGNAL PROCESSING MAGAZINE*, Marzo 2010.

- [OP06] O. Okun and H. Priisalu. Fast nonnegative matrix factorization and its application for protein fold recognition. *EURASIP J. Appl. Signal Process*, 2006.
- [Ots79] N. Otsu. A threshold selection method from gray level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979.
- [Pol07] T. Pollan. Capítulo 3 el ruido en los sistemas digitales. *Universidad de Zaragoza*, pages 288–304, 2007.
- [PP10] M. Petrou and C. Petrou. *Image Processing: The Fundamentals*. 2010.
- [PS10] A. Patil and J. Singhai. *Image denoising using curvelet transform: an approach for edge preservation*, volume 69. Journal of Scientific and Industrial Research, Enero 2010.
- [PT94] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, pages 111–126, 1994.
- [PTGM04] J. Praneet, P. Tamayo, T. R. Golub, and J. P. Merisov. Metagenes and molecular pattern discovery using matrix factorization. *PNAS*, pages 4164–4169, 2004.
- [PVL12] C. Patvardhan, A. Verma, and C. Lakshmi. Denoising of document images using discrete curvelet transform for ocr applications. *International Journal of Computer Applications*, 55(10), Octubre 2012.
- [Ram16] M. Ramos. Typewriter typeface the legacy of the writting machine in type design. 2016.
- [RMRM08] E. Ragnoli, S. McLoone, J. Ringwood, and N. Macgerailt. Matrix factorisation techniques for endpoint detection in plasma etching. In *Advanced Semiconductor Manufacturing Conference*, 2008.
- [SB03] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 177–180, 2003.
- [SCD02] J. Starck, E. Candes, and D. Donoho. The curvelet transform for image denoising. *IEEE Transaction on Image Processing*, 11(6), 2002.
- [Sch10] Reinhard Schachtner. *Extensions of Non-negative Matrix Factorization and their Application to the Analysis of Wafer Test Data*. PhD thesis, Facultad de Ciencias Naturales II - FÁsica de la Universidad de Ratisbona, Febrero 2010.
- [SDB⁺04] P. Sajda, S. Du, T.R. Brown, R.S. Stoyanova, D.C. Shungu, X. Mao, and L.C. Parra. Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain. *MedImg*, pages 1453–1465, Diciembre 2004.

- [Sha08] L. Shang. Denoising natural images based on a modified sparse coding algorithm. *Applied Mathematics and Computation*, 2008.
- [SKVR12] K. Siva, L. Koteswara, B. Vikram, and P. Ravikanth. Image compression by discrete curvelet wrapping technique with simplified sphit. *International Journal of Computer Applications*, 39(18), Febrero 2012.
- [SLWE09] B. Schuller, E. Lehmann, F. Weninger, and F. Eyben. Blind enhancement of the rhythmic and harmonic sections by nmf does it help? *Proc. of the International Conference on Acoustics, NAG/DAGA*, pages 361–364, 2009.
- [SMEG12] U. Sayed, M. Mofaddel, M. Abd Elhafiez, and M. Abdel Gawad. Image object extraction based on curvelet transform. *Applied Mathematics and Information Sciences*, 2012.
- [SS05] F. Sha and L. K. Saul. Real-time pitch determination of one or more voices by nonnegative matrix factorization. *NIPS2004*, pages 1233–1240, 2005.
- [ST09] U. Schlink and A. Thiem. Non-negative matrix factorization for the identification of patterns of atmospheric pressure and geopotential for the northern hemisphere. *International Journal of Climatology*, 2009.
- [SWH16] R. Streijl, S. Winkler, and D. Hands. Mean opinion score (mos) revisited: methods and applications, limitations and alternatives. *Multimedia System 22.2*, 2016.
- [SZCS12] L. Shang, Y. Zhou, J. Chen, and Z. Sun. Image denoising using a modified lnmf algorithm. *International Conference on Computer Science and Service System*, 2012.
- [TM13] Y. Tapia and L. Morales. Reduccion del ruido speckle en mamografias empleando filtros de difusion. *Universidad de Veracruzana, Facultad de Ingenieria Electronica y Comunicaciones*, 2013.
- [Web98] Merriam Webster. *Manual for Writers and Editors*. Merriam Webster Inc., 1998.
- [Whe95] C. Wheildon. *Type and Layout How Typography and Design Can Get Your Message Across-Or Get in the Way*. Strathmoor Pr, 1995.
- [Wit] I. Witten. Text mining. *Ciencias de la Computacion, Universidad de Waikato*.
- [XYF08] S. Xie, Z. Yang, and Y. Fu. Nonnegative matrix factorization applied to nonlinear speech and image cryptosystems. *Circuits and Systems I: Regular Papers*, Setiembre 2008.
- [YDC] L. Ying, L. Demanet, and E. Candes. 3d discrete curvelet transform. *Applied and Computational Mathematics*.

- [YDM10] A. Youssif, A. Darwish, and A. Madbouly. Adaptative algorithm for image denoising based on curvelet threshold. *International Journal of Computer Science and Network Security*, 10(1), 2010.
- [YGvV07] I. Young, J. Gerbrands, and L. van Vliet. *Fundamentals of Image Processing*, volume 2.3. Delft University of Technology, 2007.

Nomenclatura

DCR	Índice por Categorías de Degradación
FDCT	Transformada Rápida Discreta de Curvelet
IFFT	Transformada de Fourier Rápida Inversa
ITU	Unión Internacional de Telecomunicaciones
ITU-T	Sector de Normalización de las Telecomunicaciones de la ITU
KL	Kullback-Leiber
LNMF	Factorización de Matriz No Negativa Local
MAD	Desviación Absoluta Mediana
MOS	Puntuación de Opinión Media
MSE	Error Cuadrático Medio
MSSIM	Media de Semejanza Estructural
NNF	Factorización de Matrices no-Negativa
PCA	Análisis de Componente Principal
PDF	Función de Densidad de Probabilidad
PSNR	Relación Señal a Ruido Pico
SC-LNMF	Factorización de Matriz No Negativa Local de Restricción Dispersa
SP	Presentación Simultanea
SSIM	Índice de Semejanza Estructural
USFFT	Transformada de Fourier Rapida Desigualmente Espaciada